Neurociência &
Comportamento
(de verdade)



ARTIGOS CIENTÍFICOS

LETTER

doi:10.1038/nature11673

Aberrant light directly impairs mood and learning through melanopsin-expressing neurons

Tara A. LeGates[1]*, Cara M. Altimus[1]*, Hui Wang[2], Hey-Kyoung Lee[2], Sunggu Yang[2], Haiqing Zhao[1], Alfredo Kirkwood[2], E. Todd Weber[3] & Samer Hattar[1,2]

The daily solar cycle allows organisms to synchronize their circadian rhythms and sleep–wake cycles to the correct temporal niche[1]. Changes in day-length, shift-work, and transmeridian travel lead to mood alterations and cognitive function deficits[2]. Sleep deprivation and circadian disruption underlie mood and cognitive disorders associated with irregular light schedules[2]. Whether irregular light schedules directly affect mood and cognitive functions in the context of normal sleep and circadian rhythms remains unclear. Here we show, using an aberrant light cycle that neither changes the amount and architecture of sleep nor causes changes in the circadian timing system, that light directly regulates mood-related behaviours and cognitive functions in mice. Animals exposed to the aberrant light cycle maintain daily corticosterone rhythms, but the overall levels of corticosterone are increased. Despite normal circadian and sleep structures, these animals show increased depression-like behaviours and impaired hippocampal long-term potentiation and learning. Administration of the antidepressant drugs fluoxetine or desipramine restores learning in mice exposed to the aberrant light cycle, suggesting that the mood deficit precedes the learning impairments. To determine the retinal circuits underlying this impairment of mood and learning, we examined the behavioural consequences of this light cycle in animals that lack intrinsically photosensitive retinal ganglion cells. In these animals, the aberrant light cycle does not impair mood and learning, despite the presence of the conventional retinal ganglion cells and the ability of these animals to detect light for image formation. These findings demonstrate the ability of light to influence cognitive and mood functions directly through intrinsically photosensitive retinal ganglion cells.

In mammals, all light information for image formation and regulation of behaviour is detected by the retina and signalled to the relevant brain targets through retinal ganglion cells (RGCs). Most RGCs signal light to thalamic relay nuclei and then to the visual cortex for image functions. A population of intrinsically photosensitive RGCs (ipRGCs[3,4]), which predominantly signal light information for non-image-forming visual functions, expresses the photopigment melanopsin and can be distinguished from most RGCs that support image tracking and detection[5]. ipRGCs project to several hypothalamic and preoptic areas such as the suprachiasmatic nucleus (SCN), subparaventricular nucleus and ventrolateral preoptic area to regulate circadian rhythms and sleep. However, they also project to limbic regions such as the lateral habenula and the medial amygdala[4,6], highlighting a possible role in the regulation of cognitive functions.

To determine how aberrant light influences behaviour, we subjected mice to an ultradian cycle consisting of 3.5-h light and 3.5-h dark (T7). Our previous studies showed that this T7 aberrant light cycle does not affect the architecture (Supplementary Fig. 1) or the total sleep levels when compared to a 12 h:12 h light–dark (T24) control cycle[7]. We also determined whether the circadian timing system was disrupted in the T7 animals by measuring core body temperature and general activity rhythms. The T7 cycle does not cause circadian arrhythmicity in either output rhythm (Fig. 1a, b and Supplementary Fig. 2), although the circadian period is lengthened.

To determine whether the T7 cycle influences the molecular basis of the circadian clock, we measured circadian changes of a molecular clock component (PER2) in central (SCN) and peripheral (liver) tissues. Mice housed in the T7 cycle show similar rhythms and localization of PER2 expression in the SCN as in littermates housed in the T24 cycle (Fig. 1c and Supplementary Fig. 3), indicating no disruption of internal rhythmicity of the SCN pacemaker. Furthermore, *Per2* levels in the liver from mice housed in the T24 or T7 light cycle were intact and showed similar phases (Fig. 1d). Together, these data show that the T7 light cycle does not disrupt sleep or cause circadian arrhythmicity. Although the circadian timing system and sleep remain intact under the T7 light cycle, mice are exposed to light pulses at all circadian phases during the experimental time course owing to the mismatch between the imposed light cycle and the period length. Thus, the T7 cycle will allow us to determine the direct influence of aberrant light exposure on mood and cognitive functions.

The T7 model causes light to appear during the night (active) phase of the animals' cycle. To determine which brain regions respond to light presented at night, we measured expression of the transcription factor c-Fos in response to an acute light pulse. After examination of ipRGC targets that are part of or known to influence the limbic system, we found light-induced c-Fos expression in the amygdala, lateral habenula and subparaventricular nucleus (Supplementary Fig. 4). This suggests that light input particularly when presented at an aberrant time of day may influence regions of the brain involved in mood and cognitive functions.

Shorter day-length during the winter months leads to a seasonal form of depression known as seasonal affective disorder (SAD), and appropriately timed light therapy can alleviate the symptoms of SAD[8]. We thus investigated whether the T7 cycle, which exposes animals to light at inappropriate times, causes depression-like behaviours in mice by evaluating sucrose anhedonia and behavioural despair in the forced swim test (FST). Mice housed in the T7 cycle showed decreased sucrose preference, indicating an increase in depression-like behaviour (Fig. 1e and Supplementary Fig. 5). This was further supported by the FST; mice housed in the T7 cycle spent significantly more time immobile than mice housed in the T24 cycle (Fig. 1f).

An established association with depression is increased serum corticosterone levels[9]. We measured serum corticosterone in animals housed in T24 or T7 cycles at four time points across the day. Although we found an intact circadian rhythm in corticosterone with a similar phase to mice housed in the T24 cycle, the overall levels of corticosterone were increased in mice housed in the T7 light cycle (Fig. 1g).

Increases in corticosterone levels as well as anhedonia are correlated with increased stress and anxiety[10]. To assay anxiety-like behaviour, we

[1]Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, USA. [2]Department of Neuroscience, Johns Hopkins University, Baltimore, Maryland 21218, USA. [3]Department of Biology, Rider University, Lawrenceville, New Jersey 08648, USA.
*These authors contributed equally to the work.

©2012 Macmillan Publishers Limited. All rights reserved

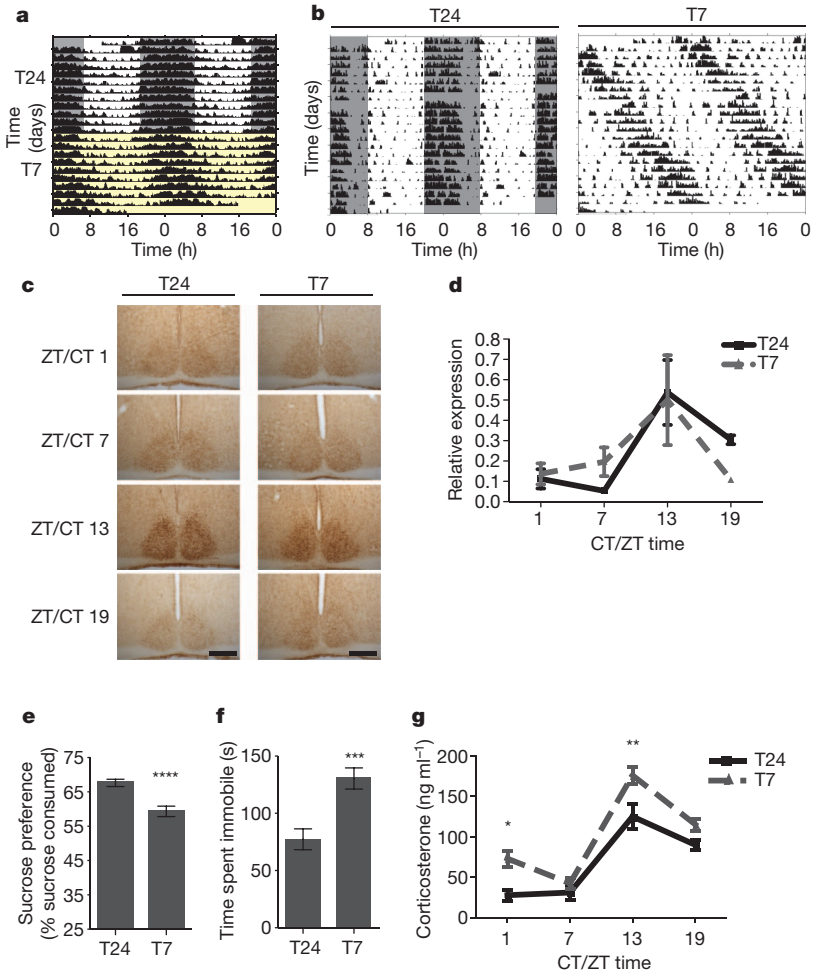


Figure 1 | Aberrant light increases depression-like behaviour and corticosterone levels. **a**, Body temperature rhythms under the T24 (grey/white) and T7 (yellow) cycles. **b**, General activity rhythms under the T24 (left) and T7 (right) cycles. **c**, PER2 expression in the SCN was rhythmic under the T24 and T7 cycles. Scale bars, 200 μm. **d**, Liver *Per2* expression was rhythmic ($P_{time} = 0.0072$) ($n = 3$–4 per time point, $P_{light\ cycle} = 0.8482$). CT, circadian time; ZT, zeitgeber time. **e**, T7 mice showed sucrose anhedonia ($n = 16$ (T24) and $n = 15$ (T7), $P < 0.0001$). **f**, T7 mice showed increased immobility in the FST ($n = 22$ per group, $P = 0.0002$). **g**, Corticosterone levels were rhythmic but increased in T7 mice ($n = 5$–6 per time point, $P_{time} < 0.0001$, $P_{light\ cycle} < 0.0001$, Bonferroni post-test: $P < 0.05$ ZT/CT 1 and $P < 0.01$ ZT/CT 13). $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$. Error bars indicate s.e.m.

used three tests: open field test, light–dark box and elevated plus maze. We found no significant difference between animals maintained in the T24 cycle and those in the T7 cycle (Supplementary Fig. 6). These results show that the T7 cycle does not globally influence behaviour but specifically elicits depression- and not anxiety-like behaviours.

Increased serum corticosterone levels and depression have been closely associated with hippocampal learning deficits[11]. We conducted a hippocampal-dependent task, the Morris water maze (MWM) (Fig. 2a–d). In the MWM, the T7 mice required significantly more trials to achieve the same latency in locating a hidden platform compared with the T24 mice (Fig. 2b), despite similar swim speeds. All our measurements were done during the day in the T24 mice (Supplementary Fig. 7). Previous studies using the MWM have shown that mice use a hippocampal-dependent spatial strategy and a hippocampal-independent non-spatial strategy to locate the platform[12]. We sought to determine whether the deficit in learning acquisition in the T7 mice was due to the lack of hippocampal-dependent spatial learning. We therefore conducted a probe trial, in which the platform in the water maze was removed from the target quadrant[13] (Fig. 2c). We found that mice housed in the T24 cycle showed a significant preference for the target quadrant, whereas mice housed in the T7 cycle showed no preference for the target quadrant (Fig. 2c). Using a reversal assay, mice housed in the T24 cycle required significantly more trials to locate the platform in the new quadrant, whereas the T7 mice showed no change in the latency to locate the platform (Fig. 2d). This further supports the dependence of T7 mice on a non-spatial escape strategy, highlighting a hippocampal- dependent learning deficit.

Spatial learning deficits are usually associated with long-term potentiation (LTP)[14] decrement in the hippocampus. We found that mice housed in both the T24 and T7 cycles have similar basal synaptic transmission at the Schaffer collaterals in the hippocampus (Fig. 2e, f). However, mice housed in a T7 cycle showed impaired LTP in response to both one and four pulses of theta burst stimulation (TBS; Fig. 2g, h). We found no difference in low-frequency stimulation-induced long-term depression (LTD) between mice housed in the T24 and T7 cycles (Fig. 2i). Impaired LTP with no change in LTD has been previously associated with sleep deprivation[15]. The selective LTP deficits observed in mice housed in the T7 aberrant light cycle independent of sleep deprivation indicates that learning impairments due to inappropriate light exposure and sleep deprivation may use similar neural pathways (see model in Supplementary Fig. 13).

To determine whether the learning deficits in the T7 cycle extend to tasks involved in hippocampal-dependent recognition memory, we conducted the novel object recognition test[16]. Mice housed in the T24 cycle showed a significant preference for the novel object (Fig. 2j). By contrast, mice housed in the T7 cycle showed no preference for a novel object from a familiar object (Fig. 2j).

©2012 Macmillan Publishers Limited. All rights reserved

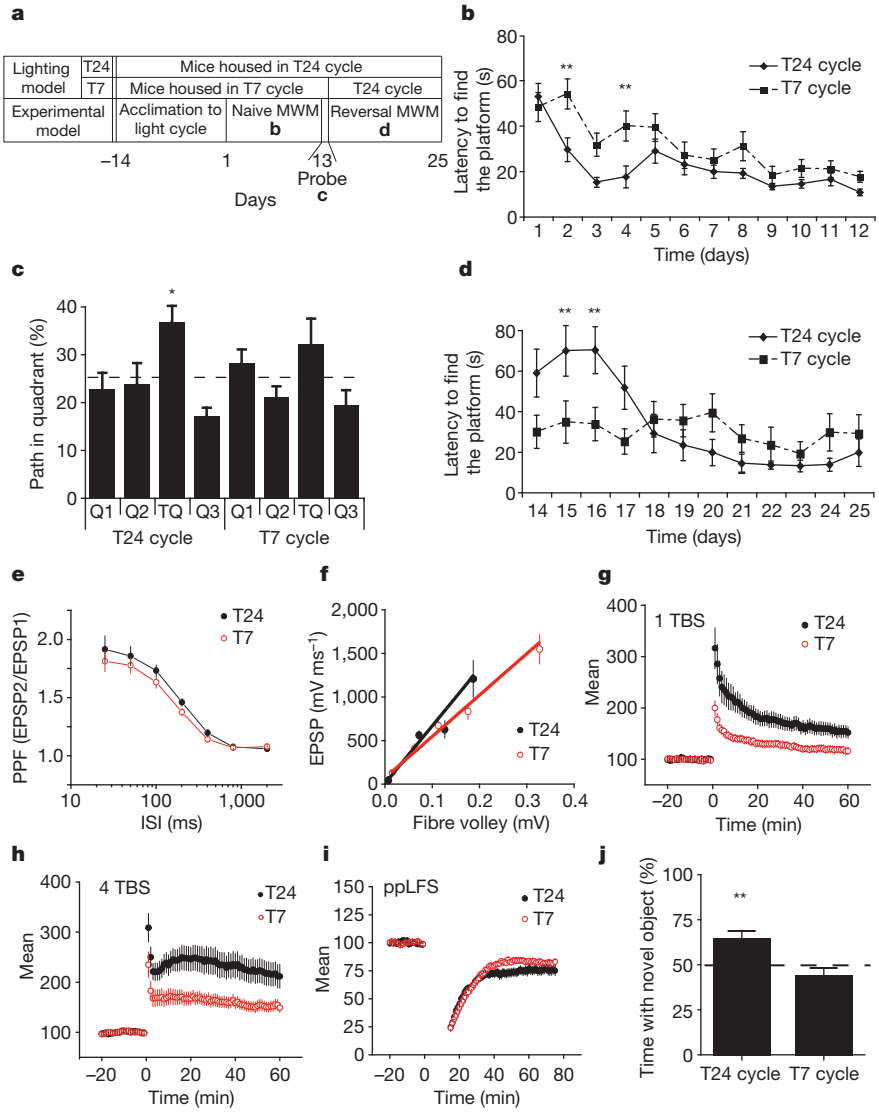


Figure 2 | Aberrant light impairs hippocampal learning, LTP, and recognition memory. **a**, Experimental model for the MWM. **b**, T7 mice showed impaired learning ($n = 21$ per group, $P_{interaction} = 0.041$, Bonferroni post-test $P < 0.01$ days 2 and 4). **c**, T7 mice show no significant preference for the target quadrant (TQ) (T24: $n = 9$, $P = 0.011$; T7: $n = 9$, $P = 0.25$). Dotted line indicates chance (25%). **d**, Mice housed in the T24 cycle showed an increased latency in a reversal trial, whereas T7-housed mice show a similar latency to the acqusition phase ($n = 9$ per group, $P_{interaction} < 0.0003$, Bonferroni post-test: $P < 0.01$ days 15 and 16). **e**, **f**, T24 and T7 mice have similar basal synaptic release properties ($P_{interaction} = 0.7256$) ($n = 5$ per group) (fibre volley: $P_{interaction} = 0.984$, slope: $P > 0.1$). EPSP; excitatory postsynaptic potential; ISI, interstimulus intervals; PPF, paired-pulse facilitation. **g**–**i**, T7 mice showed LTP deficits ($n = 5$ per group, 1 TBS: $P_{interaction} < 0.01$; 4 TBS: $P_{interaction} < 0.01$) (**g** and **h**) but no difference in LTD (**i**). ppLFS, paired-pulse low-frequency stimulation. **j**, T7 mice showed deficit in novel object recognition (T24: $n = 24$, $P = 0.0057$; T7: $n = 24$, $P = 0.1773$). $^*P < 0.05$; $^{**}P < 0.01$. Error bars indicate s.e.m.

To investigate whether antidepressants could rescue the learning deficits observed in T7 mice, we chronically administered fluoxetine to T24 and T7 mice (Fig. 3a). Chronic fluoxetine treatment decreased depression-like behaviour in mice housed in the T7 cycle (Fig. 3b). Furthermore, this treatment was able to rescue the learning deficit observed in the novel object recognition test (Fig. 3c). In support of this behavioural rescue of hippocampal function, chronic fluoxetine treatment also rescued the LTP deficit induced by the T7 cycle (Fig. 3d). Subchronic fluoxetine treatment did not rescue the increased depression or learning defect observed in mice housed in the T7 cycle (Supplementary Fig. 8), consistent with published reports[17]. We also used desipramine, a tricyclic antidepressant, and found that desipramine restored learning (Supplementary Fig. 9 and Supplementary Information). These results show that the detrimental behavioural changes induced by aberrant light exposure can be alleviated with antidepressant administration.

To determine the mechanism by which fluoxetine restores learning, we measured the circadian period in T7 mice treated with fluoxetine. Although fluoxetine can phase shift the circadian oscillator[18], chronic administration of fluoxetine did not change the length of the circadian period in mice housed in the T7 cycle (Supplementary Fig. 10). We then examined the corticosterone rhythms in fluoxetine-treated T7 mice and showed that corticosterone rhythms persist with lower overall levels compared with untreated mice (Fig. 3e). These data indicate that fluoxetine treatment does not alter circadian rhythms but instead lowers the level of corticosterone, which could lead to lower depression-like behaviour and better learning.

Studies have suggested that ipRGCs, in addition to affecting reflexive, irradiance-dependent non-image forming visual functions, might directly influence higher cognitive functions and brain processing of emotions[19–22]. To determine directly whether ipRGCs mediate the effects of the T7 cycle on mood and learning, we tested mice lacking ipRGCs (*Opn4*$^{aDTA/aDTA}$, herein referred to as aDTA mice). Although most ipRGCs are ablated, these mice still retain more than 95% of RGCs and are capable of image formation[5]. First, we compared wild

©2012 Macmillan Publishers Limited. All rights reserved

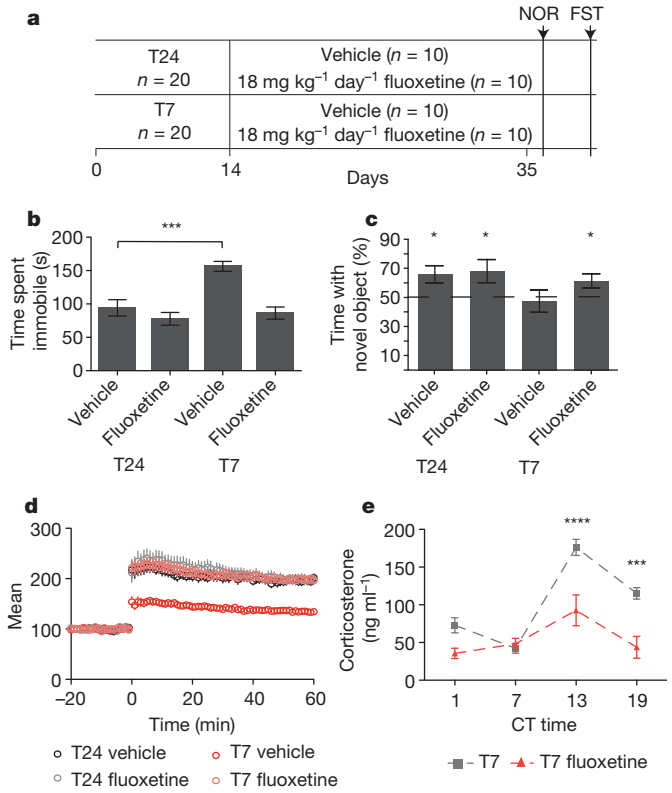


Figure 3 | Chronic antidepressant administration rescues learning. **a**, Timeline for fluoxetine administration and testing. NOR, novel object recognition. **b**, Chronic fluoxetine reduced T7 immobility time in the FST ($n = 10$ per group, $P_{interaction} = 0.0092$, Bonferoni post-test: $P < 0.001$ T24 versus T7 vehicle). **c**, T7 mice treated with fluoxetine showed restored novel object preference ($T24_{vehicle}$: $n = 9$, $P = 0.03$; $T24_{fluoxetine}$: $n = 9$, $P = 0.048$; $T7_{vehicle}$: $n = 10$, $P = 0.77$; $T7_{fluoxetine}$: $n = 10$, $P = 0.04$). **d**, Chronic fluoxetine rescued the T7-induced LTP deficit ($n = 5$ (T24 vehicle), $n = 5$ (T24 fluoxetine), $n = 5$ (T7 vehicle), and $n = 4$ (T7 fluoxetine); $P_{interaction} < 0.0001$). **e**, Chronic fluoxetine decreased corticosterone levels in T7 mice ($n = 5$–8 per time point). T7 corticosterone levels are plotted for comparison. ($P_{interaction} = 0.0053$, Bonferroni post-test: $P < 0.0001$ CT 13 and $P < 0.001$ CT 19.) $^*P < 0.05$; $^{***}P < 0.001$; $^{****}P < 0.0001$. Error bars indicate s.e.m.

type to aDTA littermates and showed no differences in baseline for the FST and corticosterone levels (Supplementary Fig. 11). We then placed aDTA mice under T24 and T7 cycles and found no difference in anxiety-like behaviours between the two groups (Supplementary Fig. 12). By contrast, the increased depression-like behaviour, learning deficits and hippocampal LTP decrement observed in the T7 cycle in wild-type animals were not observed in aDTA mice exposed to this cycle (Fig. 4). These results indicate that the negative influence of the aberrant light cycles on behaviour requires ipRGCs. Thus, subconscious light detection in humans via ipRGCs may be responsible for the depression and learning deficits observed under disruptive light environments.

Manipulation of the light environment can lead to disruptions in circadian rhythms and sleep and also cause mood and learning defects in mice and humans[2,23,24]. These studies have led to a model in which light affects cognition exclusively through the modulation of sleep and circadian pathways (Supplementary Fig. 13). Here, we provide a further model in which light directly influences mood leading to learning deficits in the context of an intact circadian timing system and normal sleep distribution and architecture (Supplementary Fig. 13a). Further support that the effects of the T7 cycle are not due to circadian disruptions originate from behavioural comparisons to animals lacking a functional circadian oscillator (for example, circadian mutant- or SCN lesioned- animals), in which reduced depression-like behaviour is observed[25,26] (Supplementary Fig. 13b). Taken together,

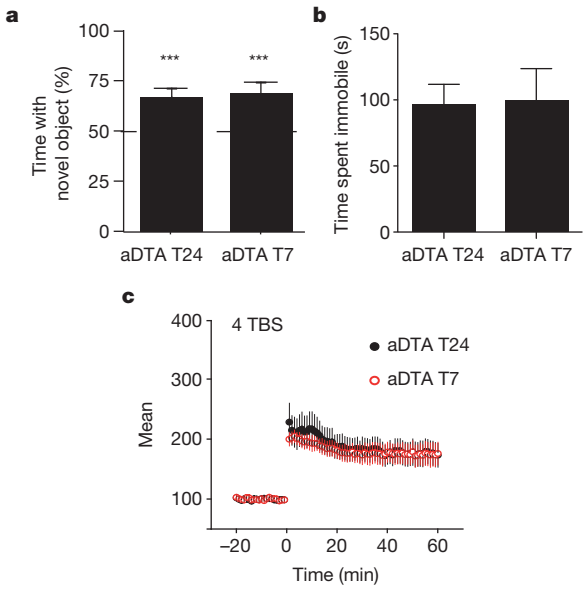


Figure 4 | ipRGCs mediate impairment of mood and learning by aberrant light. **a**, Mice lacking ipRGCs showed a significant preference for the novel object when housed in both the T24 and T7 light cycles ($n = 12$ per group, $P < 0.001$). **b**, aDTA mice housed in the T7 cycle performed similarly to those housed in the T24 cycle in the FST ($n = 6$ (T24) and $n = 5$ (T7), unpaired t-test: $P = 0.93$). **c**, T7 aDTA mice had similar LTP to T24 aDTA mice in response to four-pulse TBS. $^{***}P < 0.001$. Error bars indicate s.e.m.

these data provide evidence that the aberrant light effects on mood are direct. Furthermore, the depression-like behaviour caused by the T7 aberrant light cycle can serve as an alternative experimental model for depression research in rodents, independent of genetic manipulations or aversive treatments such as repetitive restraint or electric shocks.

METHODS SUMMARY

Adult (4–8 months) male mice (B6/129 F_1 hybrid; Jackson Laboratory) were used for all experiments involving wild-type mice placed under different light environments. Adult littermate aDTA mice placed under different light environments were of B6/129 background (6–8 months) and were raised in our laboratory. aDTA mice could not be used until they were at least 6 months old because elimination of the melanopsin cell population in these mice is not complete until this time[5]. All mice were individually housed in standard animal facility cages with access to food and water *ad libitum*. All mice were initially entrained to a 12 h:12 h light–dark cycle (T24), after which the lights for one group were switched to a 3.5 h:3.5 h light–dark cycle (T7) for 2 weeks. The light intensity during the light portion was ~800 lx, chosen to cause no circadian arrhythmicity. All experiments were done in accordance with the regulations set forth by Johns Hopkins University and Rider University Animal care and use committee.

Behavioural analysis. Mice were maintained in the T24 or T7 light cycle for 2 weeks before testing unless otherwise indicated. Body temperature and general activity were measured to evaluate circadian rhythmicity. Sucrose preference and the FST were used to assess depression-like behaviour. The MWM and novel object recognition test were used to assess hippocampal-dependent learning.

Cellular and molecular analysis. To examine the circadian timing system, immunohistochemistry and qRT–PCR were used to quantify PER2 expression in the SCN and liver, respectively. Serum corticosterone levels were quantified by ELISA. Electrophysiological recordings from the dorsal hippocampus were performed to evaluate hippocampal function including LTP and LTD.

Full Methods and any associated references are available in the online version of the paper.

Received 4 April 2011; accepted 11 October 2012.
Published online 14 November 2012.

1. Reppert, S. M. & Weaver, D. R. Coordination of circadian timing in mammals. *Nature* **418,** 935–941 (2002).
2. Foster, R. G. & Wulff, K. The rhythm of rest and excess. *Nature Rev. Neurosci.* **6,** 407–414 (2005).

©2012 Macmillan Publishers Limited. All rights reserved

3. Berson, D. M., Dunn, F. A. & Takao, M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295,** 1070–1073 (2002).
4. Hattar, S., Liao, H. W., Takao, M., Berson, D. M. & Yau, K. W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295,** 1065–1070 (2002).
5. Güler, A. D. *et al.* Melanopsin cells are the principal conduits for rod–cone input to non-image-forming vision. *Nature* **453,** 102–105 (2008).
6. Hattar, S. *et al.* Central projections of melanopsin-expressing retinal ganglion cells in the mouse. *J. Comp. Neurol.* **497,** 326–349 (2006).
7. Altimus, C. M. *et al.* Rods-cones and melanopsin detect light and dark to modulate sleep independent of image formation. *Proc. Natl Acad. Sci. USA* **105,** 19998–20003 (2008).
8. Lam, R. W. & Levitan, R. D. Pathophysiology of seasonal affective disorder: a review. *J. Psychiatry Neurosci.* **25,** 469–480 (2000).
9. Nestler, E. J. *et al.* Neurobiology of depression. *Neuron* **34,** 13–25 (2002).
10. McEwen, B. S. Protective and damaging effects of stress mediators. *N. Engl. J. Med.* **338,** 171–179 (1998).
11. Cryan, J. F. & Holmes, A. The ascent of mouse: advances in modelling human depression and anxiety. *Nature Rev. Drug Discov.* **4,** 775–790 (2005).
12. Baldi, E., Lorenzini, C. A. & Corrado, B. Task solving by procedural strategies in the Morris water maze. *Physiol. Behav.* **78,** 785–793 (2003).
13. Vorhees, C. V. & Williams, M. T. Morris water maze: procedures for assessing spatial and related forms of learning and memory. *Nature Protocols* **1,** 848–858 (2006).
14. Bliss, T. V. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361,** 31–39 (1993).
15. McDermott, C. M. *et al.* Sleep deprivation causes behavioral, synaptic, and membrane excitability alterations in hippocampal neurons. *J. Neurosci.* **23,** 9687–9695 (2003).
16. Honey, R. C., Watt, A. & Good, M. Hippocampal lesions disrupt an associative mismatch process. *J. Neurosci.* **18,** 2226–2230 (1998).
17. Dulawa, S. C., Holick, K. A., Gundersen, B. & Hen, R. Effects of chronic fluoxetine in animal models of anxiety and depression. *Neuropsychopharmacology* **29,** 1321–1330 (2004).
18. Sprouse, J., Braselton, J. & Reynolds, L. Fluoxetine modulates the circadian biological clock via phase advances of suprachiasmatic nucleus neuronal firing. *Biol. Psychiatry* **60,** 896–899 (2006).
19. Lockley, S. W. *et al.* Short-wavelength sensitivity for the direct effects of light on alertness, vigilance, and the waking electroencephalogram in humans. *Sleep* **29,** 161–168 (2006).
20. Vandewalle, G. *et al.* Spectral quality of light modulates emotional brain responses in humans. *Proc. Natl Acad. Sci. USA* **107,** 19549–19554 (2010).
21. İyilikci, O., Aydin, E. & Canbeyli, R. Blue but not red light stimulation in the dark has antidepressant effect in behavioral despair. *Behav. Brain Res.* **203,** 65–68 (2009).
22. Warthen, D. M., Wiltgen, B. J. & Provencio, I. Light enhances learned fear. *Proc. Natl Acad. Sci. USA* **108,** 13788–13793 (2011).
23. Fonken, L. K. *et al.* Influence of light at night on murine anxiety- and depressive-like responses. *Behav. Brain Res.* **205,** 349–354 (2009).
24. Ma, W. P. *et al.* Exposure to chronic constant light impairs spatial memory and influences long-term depression in rats. *Neurosci. Res.* **59,** 224–230 (2007).
25. Roybal, K. *et al.* Mania-like behavior induced by disruption of CLOCK. *Proc. Natl Acad. Sci. USA* **104,** 6406–6411 (2007).
26. Tataroğlu, O., Aksoy, A., Yilmaz, A. & Canbeyli, R. Effect of lesioning the suprachiasmatic nuclei on behavioral despair in rats. *Brain Res.* **1001,** 118–124 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank T. Gould, G. Ball and A. Sawa for their expert advice on the behavioural tests. We would like to thank R. Kuruvilla for her critical reading and advice on this manuscript. We would also like to thank the mouse tri-laboratory for suggestions and advice. This work was supported by the David and Lucile Packard Foundation grant to S.H.

Author Contributions T.A.L., C.M.A., H.Z., E.T.W. and S.H. designed experiments. T.A.L. and C.M.A. carried out experiments. H.W., H.-K.L., S.Y. and A.K. designed and performed electrophysiological experiments. T.A.L., C.M.A., H.Z., E.T.W. and S.H. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.H. (shattar@jhu.edu).

©2012 Macmillan Publishers Limited. All rights reserved

METHODS

Animals and housing. Adult (4–8 months) male mice (B6/129 F_1 hybrid; Jackson Laboratory) were used for all experiments involving wild-type mice placed under different light environments. Adult littermate aDTA mice placed under different light environment were of B6/129 background (6–8 months) and were raised in our laboratory. aDTA mice could not be used until they were at least 6 months old because elimination of the melanopsin cell population in these mice is not complete until this time[5]. All mice were individually housed in standard animal facility cages with access to food and water *ad libitum*. All mice were initially entrained to a 12 h:12 h light–dark cycle (T24), after which the lights for one group were switched to 3.5 h:3.5 h light–dark (T7) for 2 weeks. The light intensity during the light portion was ~800 lx, chosen to have no circadian arrhythmicity. All experiments were done in accordance with the regulations set forth by Johns Hopkins University and Rider University Animal care and use committee.

Body temperature measurements. Body temperature measurements were made using G2 E-mitter telemetric probes from Mini Mitter (Respironics). The telemetric probe was implanted into the peritoneum and sutured to the inside of the abdominal wall. Mice were given 1 week to recover from surgery before any recording.

Recordings were obtained using Vitalview software (Respironics), at a rate of 30 measurements per hour. Throughout, experiment mice were maintained in their home cage without perturbation and the light cycles were adjusted to test response to both the T24 and T7 light cycles. Circadian period was determined by fitting a regression line to the onsets of activity over a 7-day period using Clocklab (Actimetrics). Period lengths of mice housed in the T24 and T7 light cycles were compared using an unpaired Student's *t*-test.

General activity measurements. General activity was measured using infrared motion detectors from Mini Mitter (Respironics). Mice were housed individually, and the motion detector was mounted to the top of the cage so that general activity could be monitored throughout the T24 and T7 light cycle. Data were collected in 10-min bins using Vitalview software (Respironics). Circadian period length was determined by fitting a regression line to the onsets of activity over a 7-day period using Clocklab (Actimetrics). Period lengths of mice housed in the T24 and T7 light cycles were compared using an unpaired Student's *t*-test.

Molecular rhythm measurements. Mice were housed in either the T24 ($n = 16$) or the T7 ($n = 16$) light cycle for 2 weeks, during which general activity was monitored in T7 mice to determine their circadian phase. Mice were sampled across four time points. Mice under the T24 cycle were sampled at ZT 1, 7, 13 and 19, and mice under the T7 cycle were sampled at CT 1, 7, 13 and 19, which was determined using their general activity rhythm. Each mouse was removed from its cage, and tail blood was quickly sampled. The mouse was then anaesthetized with 1 ml of avertin (20 mg ml^{-1}). Once anaesthetized, a small sample of liver was isolated. The mouse was then perfused transcardially with 0.9% saline followed by 4% paraformaldehyde. The brain was removed and postfixed overnight in 4% paraformaldehyde followed by cryoprotection and embedding in OCT compound (Sakura Finetek).

Real-time PCR. Liver samples were homogenized and RNA extracted using the RNeasy mini kit (Qiagen). The Retroscript kit (Ambion) was used to reverse transcribe poly(A) RNAs. Real-time quantitative PCR was performed with iQ SYBR Green Supermix and the iCycler iQ real-time PCR detection system (Bio-Rad). Each sample was analysed in triplicate reactions of 50 µl. Primers for *Per2* were forward: 5′-GCCTTCAGACTCATGATGACAGA-3′ and reverse: 5′-TTTGTGTGCGTCAGCTTTGG-3′. Primers for 18S rRNA (internal control) were forward: 5′-CGCCGCTAGAGGTGAAATTC-3′ and reverse: 5′-TTGGCAAATGCTTTCGCTC-3′.

Data were analysed using the $\Delta\Delta C_t$ method, normalizing each sample to the internal control, and relative messenger RNA was determined as the percentage of the maximum value observed in the experiment. Data were analysed by two-way analysis of variance (ANOVA) followed by a Bonferroni post-hoc test to examine differences over time as well as potential light cycle effects on *Per2* expression.

PER2 immunohistochemistry. Brains were sectioned (40 µm) by cryostat through the rostral–caudal extent of the SCN, and were stored free floating in 0.1 M phosphate buffer. Free-floating sections were incubate in blocking buffer (0.1 M phosphate buffer, 3% triton, 0.5% bovine serum albumin and 1% goat serum) for 2 h. Sections were then incubated in rabbit anti-Per2 (Alpha Diagnostic International; 1:4,000 in blocking buffer) overnight and visualized with a Vectastain horseradish peroxidase kit (Vector Labs) using 3,3′-diaminobenzidine (DAB; Sigma). Sections were mounted on microscope slides, dehydrated and coverslipped with Permount. Sections were imaged at ×10 magnification with a Zeiss Axio Imager M1 microcsope. Optical density was measured using ImageJ. Data were analysed by two-way ANOVA followed by a Bonferroni post-hoc test to examine differences over time as well as potential light cycle effects on PER2 expression.

Corticosterone measurement. Serum was isolated from collected tail blood and assayed for corticosterone by ELISA (Assaypro). Data were analysed by two-way ANOVA followed by a Bonferroni post-hoc test to examine differences over time as well as light cycle effects on corticosterone levels. For fluoxetine experiments, corticosterone levels were analysed by two-way ANOVA followed by a Bonferroni post-hoc to examine differences over time as well as the effect of fluoxetine.

For the experiment comparing wild-type and aDTA mice, serum was sampled from wild-type and aDTA mice at ZT 13/CT 13 based on general activity rhythms (as described above). The samples were processed as described above. Corticosterone levels of wild-type and aDTA mice were analysed by unpaired Student's *t*-test.

Light-induced c-Fos expression. Mice were housed under a 12 h:12 h light–dark cycle and were exposed to a 10-min light pulse at ZT 14 (2 h after light offset), after which they were placed back in the dark for a further 80 min. Control mice remained in the dark until anaesthetization. Ninety minutes after the start of light presentation, mice were deeply anaesthetized with 1 ml of avertin (20 mg ml^{-1}). Once anaesthetized, the mice were perfused transcardially with 0.9% saline followed by 4% paraformaldehyde. Brains were removed, postfixed overnight in 4% paraformaldehyde, and then transferred to 0.1 M phosphate buffer. Brains were sectioned (40 µm) through the rostro–caudal extent of the SCN using a vibrotome (World Precision Instruments). Sections were stored free floating in 0.1 M phosphate buffer. Every other section was stained immunohistochemically for c-Fos. Sections were incubated in blocking buffer (0.1 M phosphate buffer, 3% Triton X-100 and 0.5% bovine serum albumin) for 2 h. Sections were incubated in rabbit anti-c-Fos (Calbiochem Ab-5; 1:20,000) overnight at 4 °C and then visualized with a goat anti-rabbit Vectastain horseradish peroxidase kit (Vector Labs) using DAB (Sigma) as a chromagen. Sections were mounted on microscope slides, dehydrated and coverslipped with Permount. Slides were viewed and imaged on a Zeiss Axio Imager.M1 microscope at ×5 magnification. Photoshop and ImageJ were used to count c-Fos-positive cells and measure the area of region counted from. The number of c-Fos positive cells was normalized to the area of the region, and these values were compared between mice that received a light pulse and dark controls. These data were analysed using an unpaired Student's *t*-test.

Sucrose anhedonia. Mice were housed in the presence of two water bottles 1 day before testing to acclimate them to the bottles. Sucrose preference was assessed over 2 days. Each day, one bottle containing 1% sucrose and one bottle containing water were introduced at the beginning of the active phase. The position of these bottles was switched 6 h later, and the bottles were removed at the end of the active phase. Bottles were weighed at the beginning and end of the active period to measure amount consumed. Sucrose preference was calculated by dividing the amount of sucrose consumed by the total amount consumed (water and sucrose). The percentage of sucrose consumed by mice in the T24 and T7 cycles was compared by a Student's *t*-test.

FST. Mice were individually placed in an inescapable container of water (25 °C) for 6 min. Behaviour was monitored by video cameras positioned in front of the apparatus and scored by a video tracking system (Forced Swim Test, Bioserve). Time spent immobile for the last 4 min of the test was calculated. Increased time spent immobile is indicative of increased depression-related behaviour. The amount of time spent immobile during the last 4 min was analysed by Student's *t*-test. For fluoxetine experiments, the amount of time spent immobile during the last 4 min were analysed by two-way ANOVA followed by a Bonferroni post-hoc to compare light cycle and drug treatment.

Open field. Mice were individually placed in the centre of a large, brightly lit arena (500 lx, 60 × 60 cm) and allowed to explore for 5 min. Behaviour was monitored from above by a video camera connected to a computerized video tracking system (Anymaze, Stoelting). The apparatus was cleaned thoroughly between each trial. The percentage distance travelled and time in the centre of the arena were measured. These measures were compared between mice housed in the T24 and T7 cycles using an unpaired Student's *t*-test.

Light–dark box. The light–dark box consisted of two compartments equivalent in size (20 × 20 cm); one area is brightly lit (600 lx) and the other is dimly lit (<1 lx). A small opening joins the two compartments, so the mice could freely move between the two areas. Mice were dark adapted for 1 h before testing. At the beginning of the test, dark-adapted mice were placed in the lit compartment facing away from the opening and allowed to freely explore for 5 min. Behaviour was monitored from above by a video camera connected to a computerized video tracking system (Anymaze). The apparatus was cleaned thoroughly between each trial. The number of transitions between the two compartments as well as the time spent and distance travelled in the lit room were measured as indications of anxiety-related behaviour. These measures were compared between mice housed in the T24 and T7 cycles using an unpaired Student's *t*-test.

Elevated plus maze. The apparatus consisted of two open arms (42 × 6 cm) opposite to one another and two arms enclosed by walls (42 × 6 × 14 cm) opposite

©2012 Macmillan Publishers Limited. All rights reserved

of one another forming a cross. The arms were separated by a central platform (6 × 6 cm). The maze was elevated (33 cm) such that the open arms convey openness, unfamiliarity and elevation. The light intensity in the open arms was ~600 lx, whereas the light intensity in the closed arms was ~200 lx. Mice were placed in the centre of the elevated plus maze facing one of the open arms. Behaviour was monitored from above by a video camera connected to a computerized video tracking system (Anymaze). The apparatus was cleaned thoroughly between each trial. The time spent and the distance travelled in the open arms were measured as indications of anxiety-related behaviour. These measures were compared between mice housed in the T24 and T7 cycles using an unpaired Student's *t*-test.

MWM. The water maze consisted of a circular pool (150 cm in diameter) with room temperature water (26–28 °C). The water was made opaque with the addition of non-toxic white tempura paint to hide the escape platform. The platform was made from PVC piping with a top (10 cm in diameter) painted white and submerged in the pool such that 1 cm of water covered the platform hiding it from sight. For visual trials, a flag made from a 50-ml conical tube covered with coloured tape was placed on the platform. During the acquisition, probe and reversal trials, four cues were attached to the side of the pool equidistant from one another, and the entire pool was surrounded by a plain curtain to block any other visual cues. Performance was scored using a video tracking system (Anymaze) with a camera mounted above the pool. The light intensity at the water surface was approximately 500 lx.

Mice were tested in four stages: visual, acquisition, probe and reversal (see Fig. 2a). Before spatial training, mice were trained to escape the pool using the visual flag located on top of the platform to familiarize them with the test. This was performed four times with an inter-trial interval of (30 min), and the platform was moved between each trial. We also used this visual training to screen for and remove animals that show floating behaviour in the water maze. Animals that floated in two or more trials during training were not tested in the spatial task to prevent confounds from floating behaviour. During the acquisition phase, mice were trained (one trial per day for 12 days) to find the hidden platform using the four visual distal cues surrounding the pool. The mouse was randomly placed in a different area of the pool at the start of each trial with the platform maintained in the same quadrant (target quadrant). The platform was removed on day 13 in the probe trial. The swimming in each quadrant and specifically the preference for the target quadrant was measured to evaluate spatial memory using a computerized video tracking system (Anymaze). Reversal training began on day 14 when the platform was moved to the quadrant opposite the original target quadrant. Mice were trained as described for the acquisition phase.

Latency to locate the platform during the acquisition and reversal phases was analysed by two-way ANOVA followed by a Bonferroni post-hoc test to examine changes in latency throughout the course of the experiment as well as the effect of light cycle exposure. Probe trial was analysed by calculating the percentage time spent in the target quadrant and performing a one-sample *t*-test to determine whether this was significantly above 25%.

Novel object recognition. Novel object recognition was composed of three stages: acclimation to the novel object arena, familiar object exposure and finally novel object exposure. Mice were first removed from their home cage, acclimated to the empty testing arena (light intensity of 500 lx) for 10 min, and subsequently returned to their home cage 24 h before including two objects in the arena. The day after acclimation, mice were returned to this arena with two identical objects that they could freely explore for 10 min, after which they were returned to their home cage for 1 h. At the end of the 1-h period, mice were placed back into the arena with one of the objects changed to a novel object, and were allowed to explore both the familiar and novel objects for 5 min. Behaviour was monitored from above by a video camera connected to a computerized video tracking system (Anymaze), and the percentage of time spent with each object was calculated. Wild-type mice spend more time with the novel object, however, mice with recognition memory deficits will not be able to distinguish the novel from the stable object. Objects had been previously tested to ensure that animals showed no initial preference for a particular object. The identity of the objects (familiar versus novel) was counterbalanced. The objects and arena were thoroughly cleaned between each trial to remove odour cues. Object recognition was analysed by calculating the percentage time spent with the novel object and performing a one sample *t*-test to determine whether this was significantly above 50%.

Slice electrophysiology. Coronal (0.4 mm) hippocampal slices were prepared as described[27] in ice-cold dissection buffer (2.6 mM KCl, 1.23 mM NaH_2PO_4, 26 mM $NaHCO_3$, 212.7 mM sucrose, 10 mM dextrose, 3 mM $MgCl_2$ and 1 mM $CaCl_2$, bubbled with 5% CO_2, 95% O_2). Recordings were done in a similar buffer but with the sucrose replaced by NaCl and the temperature raised to 30 °C. Synaptic responses were evoked at 0.33 Hz stimulating the Schaffer collaterals with 0.2-ms pulses (concentric bipolar electrodes, FHC), and recorded extracellularly in CA1 stratum radiatum. LTP was induced by TBS, consisting of one or four theta epochs delivered at 0.1 Hz. Each epoch, in turn, consisted of 10 trains of four pulses (at 100 Hz) delivered at 5 Hz. LTD was induced by low frequency stimulation (1 Hz, 15 min). These protocols were delivered after 20 min of stable baseline transmission. All hippocampal slice electrophysiological recordings were performed and analysed by an experimenter blind to the treatment of the animals. Two-way ANOVA was used to analyse fibre volley differences between the T7 and T24 treatments. The slopes for the linear fit of the fibre-volley–slope relationship were compared by *t*-test.

Fluoxetine administration. Mice were housed in either the T24 or T7 cycle for two weeks ($n = 20$ per group) with food and water *ad libitum*. For chronic treatment, this was followed by a 3-week treatment of 18 mg kg^{-1} day^{-1} of fluoxetine (Sigma). For subchronic treatment, this was followed by a 4-day treatment of 18 mg kg^{-1} day^{-1}. Fluoxetine was administered in the drinking water and control mice received tap water. Dosage was calculated based on the average amount of water consumed per day and mouse weight. Fluoxetine consumption was also measured during treatment to determine the amount consumed per mouse.

Period length measurement with fluoxetine treatment. Mice were housed under infrared motion detectors as described above. Mice were housed in the T7 light cycle for 2 weeks after which fluoxetine (18 mg kg^{-1} day^{-1}) was administered in the drinking water for 3 weeks. Circadian period was determined by fitting a regression line to the onsets of activity over a 7-day period. Period lengths before and after 3 weeks of fluoxetine treatment were compared by paired *t*-test.

Desipramine administration. Desipramine (Sigma) was dissolved in sterile water with 5% Tween-20. Each mouse received 16 mg kg^{-1} desipramine intraperitoneally 24- and 1-h before testing in the novel object recognition paradigm. The same volume of vehicle (water with 5% Tween-20) was administered intraperitoneally to control mice.

Statistical analysis. All statistical analysis was performed using GraphPad Prism. Specific tests used to analyse data are described their respective section of the methods.

27. Lee, H. K., Min, S. S., Gallagher, M. & Kirkwood, A. NMDA receptor-independent long-term depression correlates with successful aging in rats. *Nature Neurosci.* **8,** 1657–1659 (2005).

©2012 Macmillan Publishers Limited. All rights reserved

Current Biology 21, 2109–2114, December 20, 2011 ©2011 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2011.11.018

Report

Acquiring "the Knowledge" of London's Layout Drives Structural Brain Changes

Katherine Woollett[1] and Eleanor A. Maguire[1,*]
[1]Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, UK

Summary

The last decade has seen a burgeoning of reports associating brain structure with specific skills and traits (e.g., [1–8]). Although these cross-sectional studies are informative, cause and effect are impossible to establish without longitudinal investigation of the same individuals before and after an intervention. Several longitudinal studies have been conducted (e.g., [9–18]); some involved children or young adults, potentially conflating brain development with learning, most were restricted to the motor domain, and all concerned relatively short timescales (weeks or months). Here, by contrast, we utilized a unique opportunity to study average-IQ adults operating in the real world as they learned, over four years, the complex layout of London's streets while training to become licensed taxi drivers. In those who qualified, acquisition of an internal spatial representation of London was associated with a selective increase in gray matter (GM) volume in their posterior hippocampi and concomitant changes to their memory profile. No structural brain changes were observed in trainees who failed to qualify or control participants. We conclude that specific, enduring, structural brain changes in adult humans can be induced by biologically relevant behaviors engaging higher cognitive functions such as spatial memory, with significance for the "nature versus nurture" debate.

Results and Discussion

In order to qualify as a licensed London taxi driver, a trainee must learn the complex and irregular layout of London's ~25,000 streets (Figure 1) within a 6-mile radius of Charing Cross train station, along with the locations of thousands of places of interest. This spatial learning is known as acquiring "the Knowledge" and typically takes between 3 and 4 years, leading to a stringent set of examinations, called "appearances," which must be passed in order to obtain an operating license from the Public Carriage Office (PCO, the official London taxi-licensing body). This comprehensive training and qualification procedure is unique among taxi drivers anywhere in the world. Previous cross-sectional studies of qualified London taxi drivers documented more gray matter (GM) volume in their posterior hippocampi and less in their anterior hippocampi relative to non-taxi-driving matched control participants [2–4]. Moreover, correlation of hippocampal GM volume with years of taxi driving suggested that structural differences may have been acquired through the experience of navigating, to accommodate the internal representation of London, and were not merely due to preexisting hippocampal GM volume patterns disposing individuals to being taxi drivers [2–4, 19, 20]. As well as displaying a specific pattern of hippocampal GM volume, qualified taxi drivers have been found to display better memory for London-based information, but surprisingly poorer learning and memory for certain types of new visual information (e.g., delayed recall of complex figures), compared with control participants, suggesting there might be a price to pay for the acquisition of their spatial knowledge, perhaps linked to their reduced anterior hippocampal volume (see [3, 4, 20] for more on this). Interestingly, the opposite pattern of hippocampal GM volume and memory profile has been described in retired taxi drivers, hinting that any changes acquired through learning might be reversed or "normalized" when the call on stored memory representations lessens [21].

London taxi drivers are therefore a useful model of memory, illuminating the role of the hippocampus and intrahippocampal functional differentiation and potentially informing about whether hippocampal structure and memory capacity are hardwired or amenable to change. Given current economic imperatives and increasing longevity, the need to keep retraining and learning throughout adulthood has never been more acute. Direct evidence for hippocampal plasticity in response to environmental stimulation could allow us to understand the boundaries within which human memory operates and the scope for improving or rehabilitating memory in educational and clinical contexts. Moreover, given the dearth of longitudinal magnetic resonance imaging (MRI) structural association studies focusing on higher cognitive functions in average adults engaged in truly naturalistic behaviors, taxi drivers could contribute new information to the wider debate about whether key aspects of cognition are fixed or malleable.

The above aspirations are predicated upon hippocampal plasticity in adult humans, evidence for which remains sparse [10, 18]. We therefore conducted a longitudinal study examining 79 male trainee London taxi drivers at the start of their training (T1, time 1) and then again 3–4 years later just after qualification (T2, time 2), as well as 31 male control participants, with two aims. First, given that the PCO suggests that typically 50%–60% of trainees fail to qualify, we anticipated having three groups of participants: trainees who qualified (Q), trainees who failed to qualify (F), and the controls (C). With this design, we could retrospectively examine whether MRI and/or neuropsychological findings at T1 could predict who would eventually qualify 3–4 years later at T2. Second, we sought to ascertain whether the pattern of hippocampal GM volume and memory profile observed in previous cross-sectional taxi driver studies would be induced and observable within the same participants as a result of acquiring "the Knowledge."

Of the 79 trainees, 39 went on to qualify as licensed London taxi drivers, while 20 did not qualify (ceased training or failed their appearances) but agreed to return for testing at T2. Of the other 20 trainees who did not wish to return at T2, two had qualified, and the remaining 18 decided to stop training or made no appearances. Thus in our sample, 51.9% of trainees qualified, in line with PCO figures. All 31 control participants returned for testing at T2.

*Correspondence: e.maguire@ucl.ac.uk

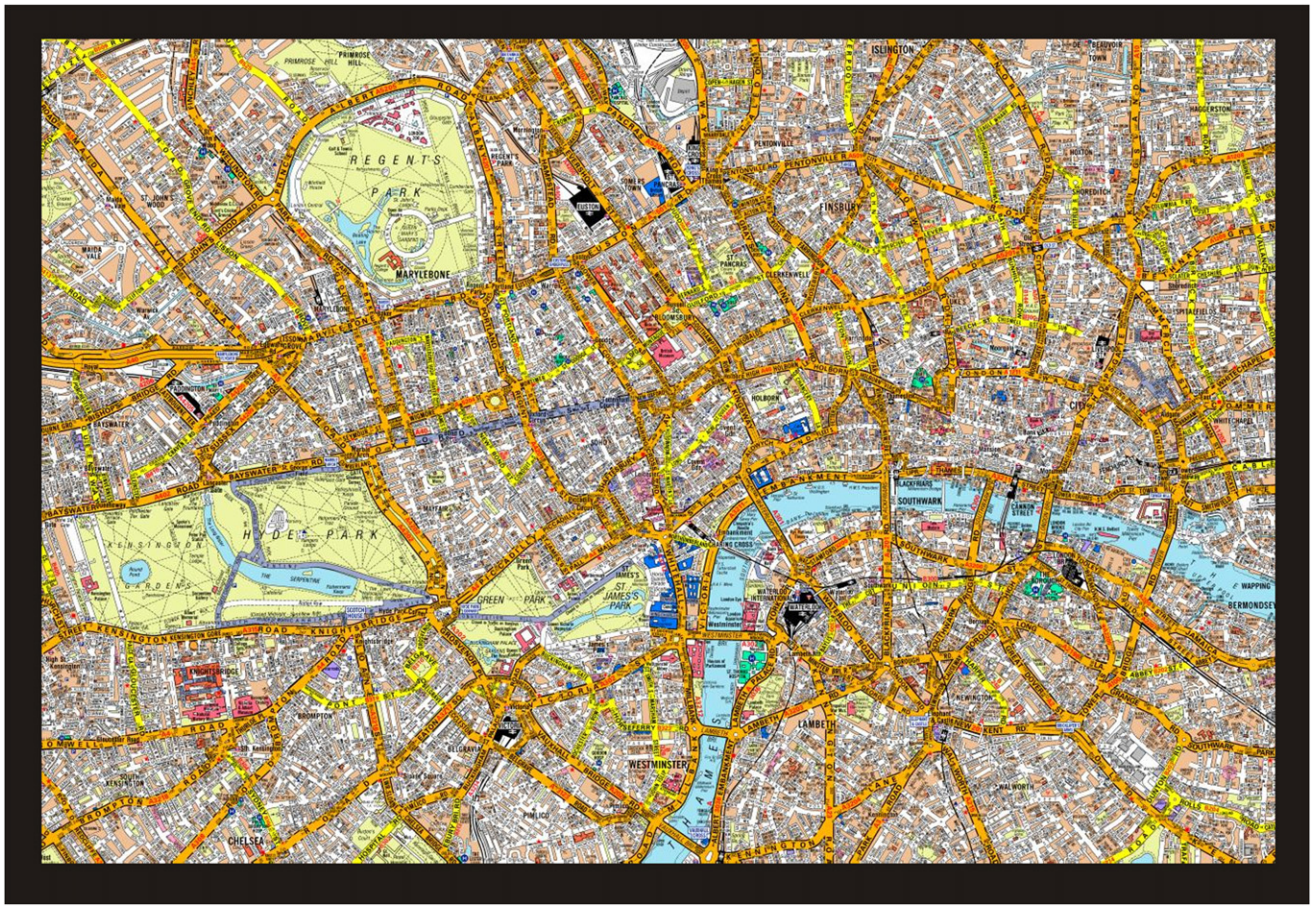


Figure 1. Central London

Trainee taxi drivers must acquire "the Knowledge" of London's layout within a 6-mile radius of Charing Cross train station. This map shows just a part of the total area that must be learned. Map reproduced by permission of Geographers' A-Z Map Co. Ltd. © Crown copyright 2005. All rights reserved. License number 100017302.

Time 1

We first focused on the data acquired at T1 and examined the status before training occurred of those trainees who subsequently qualified or not, as well as the control participants. There were no statistically significant differences between the three groups on a range of background measures such as age, handedness, education, and IQ [$F(8,166) = 1.81$; $p = 0.07$; see Table 1]. We next looked at their performance at T1 on a battery of tests assessing working and long-term memory, recognition and recall, and visual and verbal material (see Supplemental Experimental Procedures and Table S1A available online). There were no statistically significant differences between the three groups for any of the memory measures [$F(18,158) = 1.33$; $p = 0.18$]. Finally, we examined the structural MRI brain scans of participants acquired at T1, before training. An automated whole-brain analysis method, voxel-based morphometry (VBM; [22, 23]) implemented in SPM8, was used to compare GM volume between the groups. No significant differences were found, even with a liberal statistical threshold ($p < 0.005$ uncorrected for multiple comparisons).

Our data therefore showed that at T1, before acquisition of a detailed internal spatial representation of London's layout, there was no general intellectual, mnemonic, or structural brain difference that was associated with subsequent success at qualification. In particular, hippocampal GM volume, both anteriorly and posteriorly, was indistinguishable between those who would qualify, those who would not, and the controls. This means that the groups started out on equal terms and that any changes that subsequently emerged would be due to acquiring "the Knowledge."

Time 2

In the first instance, we compared the two trainee groups on a range of training-related variables (see Table 1). There was no significant difference between those who qualified and those who did not in terms of the total time they had spent in training [$t(57) = 0.67$; $p = 0.5$]. There was a difference, however, in the number of hours per week the groups spent training, with the trainees who qualified spending twice as many hours per week training compared to the group who failed to qualify [$t(57) = 6.62$; $p = 0.001$]. Unsurprisingly, the two groups differed also in the number of appearances that they made, with the failed candidates making hardly any [$t(57) = 13.1$; $p = 0.001$]. Finally, there was no significant difference in the time elapsed between testing at T1 and testing at T2 for the Q, F, and C groups [$F(2, 86) = 1.27$; $p = 0.28$].

We next examined performance at T2 on parallel versions of the memory tests that had first been employed at T1 (see Table S1B). Whereas at T1 there were no differences between

Table 1. Background Characteristics of the Participants

Measure	Qualified (n = 39)	Failed to Qualify (n = 20)	Controls (n = 31)
Age (years)	37.97 ± 7.96	40.50 ± 5.27	35 ± 8.99
Age left school (years)	16.66 ± 1.32	16.75 ± 1.40	16.77 ± 1.30
Estimated verbal IQ	97.72 ± 6.29	98.66 ± 3.49	100.79 ± 3.79
Matrix reasoning (scaled score)	11.89 ± 2.06	12.20 ± 1.98	11.83 ± 2.39
Handedness (laterality index)	87.97 ± 24.09	88.10 ± 21.50	71.74 ± 38.69
Total training time (months)	38.84 ± 7.02	35.80 ± 10.32	–
Training time per week (hours)[a]	34.56 ± 12.40	16.70 ± 8.21	–
Number of appearances[a]	15.64 ± 3.66	2.60 ± 3.45	–
Time between T1 and T2 testing (months)	35.28 ± 8.19	36.15 ± 8.53	32.8 ± 7.37

Measurements are given in means ± SD.
[a] Significantly more for trainees who qualified compared to trainees who did not qualify.

the groups, at T2 this was no longer the case: significant differences between the three groups now emerged [$F(18,158) = 2.35$; $p = 0.004$], driven by two main effects. A significant difference was found between the groups on the London landmarks proximity judgments test [$F(2,87) = 8.09$ $p = 0.001$], with trainees who qualified being significantly better at judging the spatial relations between London landmarks than the control participants ($p = 0.003$), as were nonqualified trainees ($p = 0.003$). The groups also differed on the delayed recall of the Taylor complex figure [$F(2,87) = 4.38$; $p = 0.015$], with qualified trainees being significantly worse at recalling the complex figure after 30 min delay than the control participants ($p = 0.01$). By contrast, the performance of the nonqualified trainees was not significantly different from that of control participants ($p = 0.14$).

The memory profile displayed by the now qualified trainees mirrors exactly the pattern displayed in several previous cross-sectional studies of licensed London taxi drivers [3, 4, 20] (and that which normalized in the retired taxi drivers [21]). In those studies also, the taxi drivers displayed more knowledge of the spatial relationships between landmarks in London, unsurprisingly, given their increased exposure to the city compared to control participants. By contrast, this enhanced spatial representation of the city was accompanied by poorer performance on a complex figure test, a visuospatial task designed to assess the free recall of visual material after 30 min. Our findings therefore not only replicate those of previous cross-sectional studies but extend them by showing the change in memory profile within the same participants. That the only major difference between T1 and T2 was acquiring "the Knowledge" strongly suggests that this is what induced the memory change.

We then turned to the structural MRI brain scans acquired at T2 and compared them to the scans acquired in the same individuals at T1 in order to assess whether acquiring "the Knowledge" had any impact on GM volume. This was accomplished by implementing high-dimensional warping (HDW) in SPM8. HDW safeguards against nonspecific subtle differences that may arise between the first and second scans within subjects in a longitudinal study (see Supplemental Experimental Procedures and also [24] for full details). The time between scans, total GM volume, and participant age were modeled as confounding variables. Given our a priori interest in the hippocampus, the significance level was set at $p < 0.001$ corrected for the volume of the hippocampus; otherwise the significance level was set at $p < 0.05$ corrected for multiple comparisons across the whole brain. We first looked at the scans of the qualified trainees and found that GM volume had increased in the posterior hippocampi bilaterally (30, −42, 1, $z = 5.44$; 20, −37, 10, $z = 5.64$; 24, −39, 7, $z = 4.02$; −29, −42, 3, $z = 5.91$) at T2 relative to T1 (see Figure 2 and Figure 3). The opposite comparison (T1 > T2) did not show any significant differences, nor were there any significant differences in GM volume anywhere else in the brain for either contrast. Similar analyses were performed for the trainees who failed to qualify. No significant differences in GM volume were found anywhere in the brain, including the hippocampi, for T2 > T1 or T1 > T2. When the data from the two time points of the control participants were compared, again no significant differences in GM volume emerged.

As with the memory data, the structural MRI brain data replicate and extend the cross-sectional taxi driver studies [2–4] by showing that the increased posterior hippocampal GM volume previously observed most likely occurred as a result of acquiring the detailed spatial representation of London's layout. Importantly, the posterior hippocampal change in GM volume cannot be attributed to the qualified trainee taxi drivers' training procedures, methods, or general attempts to learn, as the trainees who failed to qualify were exposed to the same milieu. Although the posterior hippocampal increase accords with previous findings, we did not observe a decrease in anterior hippocampal GM volume. This is interesting, because it may provide an insight into the time frame of the hippocampal structural changes. It could be that they occur serially, with the increase in posterior hippocampus happening first and within 3–4 years. We speculate that the anterior hippocampal GM volume might then decrease subsequently and in response to the posterior increase. In fact, the poor performance of the qualified trainees on the delayed recall of the complex figure at T2 may be an indication that changes are already afoot in the anterior hippocampus but are not yet detectable with MRI.

That acquiring "the Knowledge," which encompasses spatial learning and memory, can drive changes in posterior hippocampus illustrates the close relationship between this region and spatial navigation [25–27] and suggests that the hippocampus acts as a storage site for the spatial information acquired during "the Knowledge," or as a processing hub for detailed navigational information. Our results underline the existence of functional differentiation in the hippocampus, with anterior and posterior regions diverging in their response to spatial memory [28, 29]. The hippocampal plasticity we have observed in vivo in adult humans parallels the effects reported in nonhumans where intraindividual hippocampal volume changes occur in response to demands placed on spatial memory [30, 31]. Having documented this hippocampal change, the question is what mechanism underpins this process.

Using standard structural MRI scanning in humans, it is not possible to address this question directly, but based on work in nonhumans, there are several candidates. Studies in rodents have demonstrated that when learning requires cognitive effort and where learning actually takes place (i.e., where material is remembered after a delay), there is an effect on the rate of hippocampal neurogenesis [32]. Moreover, the animals that learn best have more new neurons after training

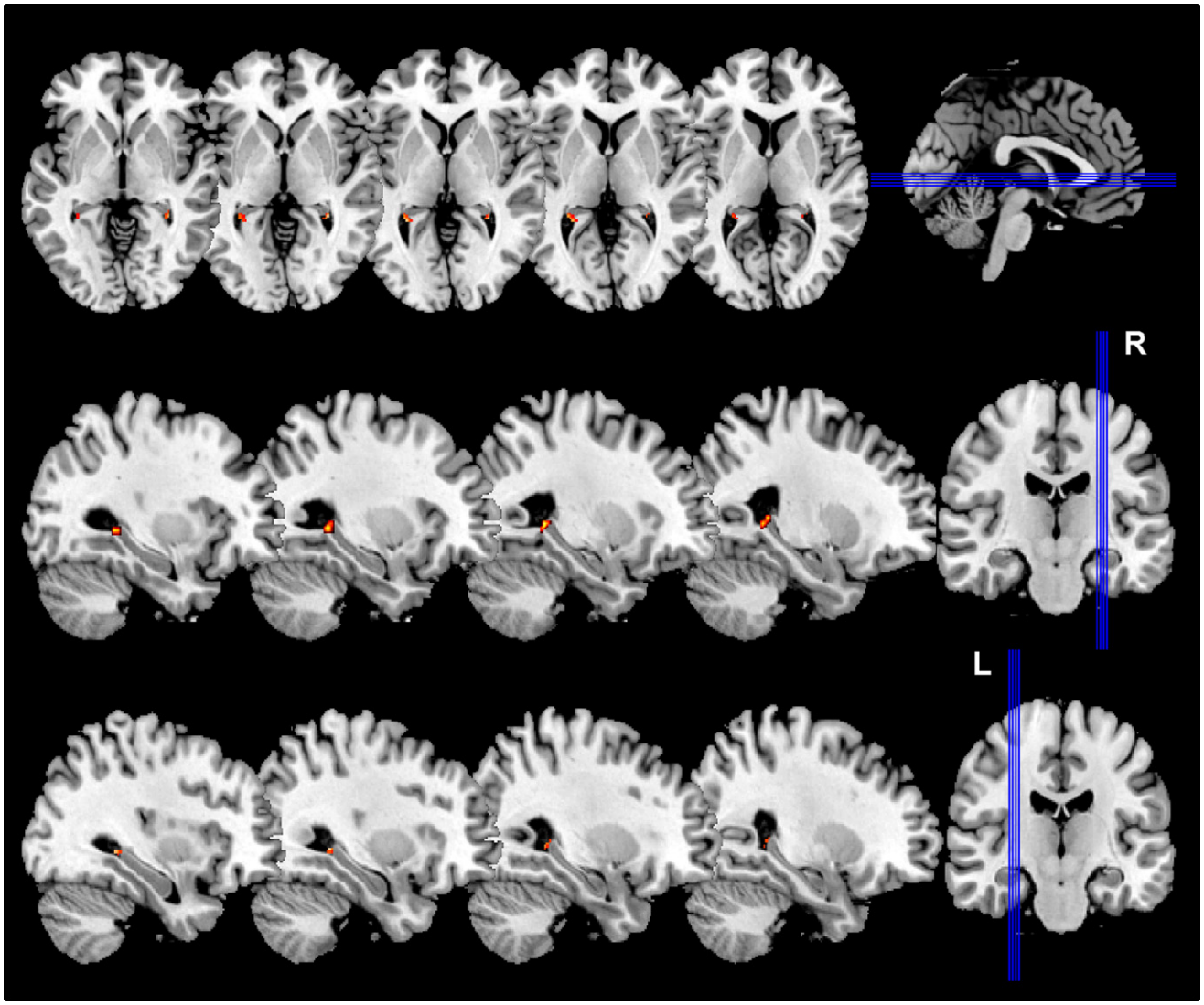


Figure 2. Gray Matter Volume Changes between T1 and T2 in Qualified Trainees

Gray matter (GM) volume increased in the most posterior part of the hippocampus bilaterally between T1 and T2 as a result of acquiring a detailed representation of London's layout. This change was only apparent in the trainees who qualified. The upper row shows axial views, and the middle and lower rows show sagittal views through the right (R) and left (L) sides of the brain, respectively, that encompass the GM changes (shown in orange and yellow) in the peak voxels detailed in the text.

than those who do not learn, or do not learn efficiently [33, 34]. If neurogenesis is what underpins the hippocampal volume change in qualified taxi drivers, it may be related to recruitment of new neurons following neurogenesis [35, 36] that are pressed into the service of spatial memory. The development of greater communication between neurons in the form of increased synaptogenesis [37, 38] might also be involved, and proliferation in dendritic arborization, augmenting connectivity between neurons, could in turn increase memory capacity and also lead to volumetric changes [39]. Glial cells, which continue to be produced, albeit at a slow pace, throughout adulthood [40], could also be implicated and have been shown to increase in volume with the addition of synapses following learning [41]. In the future, new approaches to human brain scanning in vivo may eventually be able to provide more direct insight into the key mechanisms supporting human hippocampal plasticity [42].

To conclude, we have shown that there is a capacity for memory improvement and concomitant structural changes to occur in the human brain well into adulthood. That there are many thousands of licensed London taxi drivers shows that acquisition of "the Knowledge," and presumably the brain changes that arise from it, is not uncommon, offering encouragement for lifelong learning, and possibly a role in neurorehabilitation in the clinical context. However, this needs to be balanced by our finding that memory improvement in one domain may come at the expense of memory performance elsewhere. One final point to consider concerns the PCO figures and our data showing that only half of trainees who undertake "the Knowledge" actually qualify. The reasons we were given for ceasing training included the time commitment being too great, financial imperatives, and family obligations. Very few trainees reported ceasing because they found the spatial memory demands to be too great, although it is possible, or

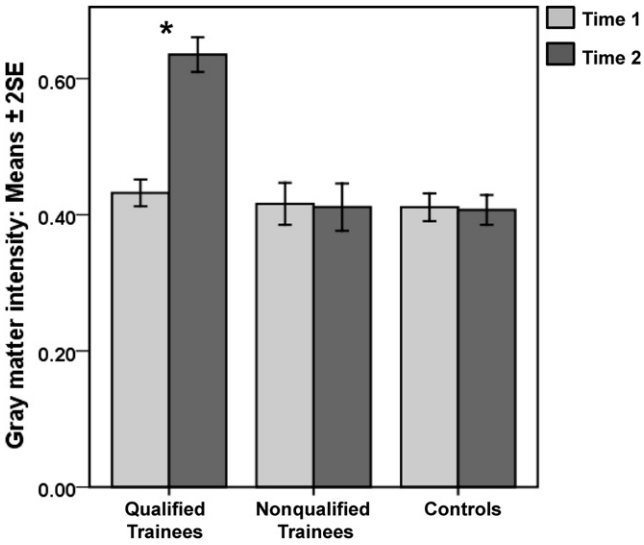


Figure 3. Plot of Gray Matter Intensities across Groups and Time

GM intensity values were extracted from the four peak hippocampal voxels identified by the HDW procedure (see text for details). Here we show a plot from one of those peaks (the others resulted in very similar plots) at 20, −37, 10 in the right posterior hippocampus. Only the qualified trainees experienced a significant increase in posterior hippocampal gray matter between T1 and T2; this change was evident in every qualified trainee. Data are presented as means ± two standard errors of the mean. *$p < 0.05$.

even likely, that the reasons given may have masked such difficulties in some individuals. It could be that there are inherited factors that feed into individual differences in spatial memory and navigation ability. One source of influence may come from genes. Genetic association studies have demonstrated effects of specific gene polymorphisms on the volume of the hippocampus [43, 44] and memory performance [45, 46]. Although our data show that environmental stimulation can drive structural brain changes, it may be that this hippocampal plasticity expresses itself only in certain individuals. The trainees that qualified may have had a genetic predisposition toward plasticity that the nonqualified individuals lacked. Thus, the perennial question of "nature versus nurture" is still open for future investigations that incorporate genetic and other influences on individual differences, as well as cognitive and structural brain dimensions.

Experimental Procedures

All participants gave informed written consent to participation in accordance with the local research ethics committee. Full details of the participants, cognitive tests, MRI scans, and data analysis procedures are provided in the Supplemental Experimental Procedures.

Supplemental Information

Supplemental Information includes one table and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.cub.2011.11.018.

Acknowledgments

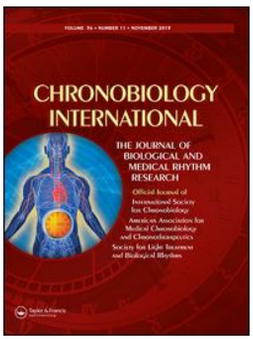
This work was supported by the Wellcome Trust. We thank all participants for their time and continued interest throughout the study. We are very grateful to the Public Carriage Office, in particular Tony Bishop and Alison Said. Thanks also to the Knowledge Schools, the Physics and Imaging Support teams at the Wellcome Trust Centre for Neuroimaging, Ralf Deichmann, Amanda Brennan, John Ashburner, Jenny Crinion, Alex Leff, Ric Davis, and Chris Freemantle.

Received: September 8, 2011
Revised: October 26, 2011
Accepted: November 8, 2011
Published online: December 8, 2011

References

1. Gaser, C., and Schlaug, G. (2003). Gray matter differences between musicians and nonmusicians. Ann. N Y Acad. Sci. *999*, 514–517.
2. Maguire, E.A., Gadian, D.G., Johnsrude, I.S., Good, C.D., Ashburner, J., Frackowiak, R.S.J., and Frith, C.D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. Proc. Natl. Acad. Sci. USA *97*, 4398–4403.
3. Maguire, E.A., Woollett, K., and Spiers, H.J. (2006). London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. Hippocampus *16*, 1091–1101.
4. Woollett, K., and Maguire, E.A. (2009). Navigational expertise may compromise anterograde associative memory. Neuropsychologia *47*, 1088–1095.
5. Mechelli, A., Crinion, J.T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R.S., and Price, C.J. (2004). Neurolinguistics: structural plasticity in the bilingual brain. Nature *431*, 757.
6. Hüfner, K., Binetti, C., Hamilton, D.A., Stephan, T., Flanagin, V.L., Linn, J., Labudda, K., Markowitsch, H., Glasauer, S., Jahn, K., et al. (2011). Structural and functional plasticity of the hippocampal formation in professional dancers and slackliners. Hippocampus *21*, 855–865.
7. Martin, S.B., Covell, D.J., Joseph, J.E., Chebrolu, H., Smith, C.D., Kelly, T.H., Jiang, Y., and Gold, B.T. (2007). Human experience seeking correlates with hippocampus volume: convergent evidence from manual tracing and voxel-based morphometry. Neuropsychologia *45*, 2874–2881.
8. Kanai, R., Feilden, T., Firth, C., and Rees, G. (2011). Political orientations are correlated with brain structure in young adults. Curr. Biol. *21*, 677–680.
9. Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., and May, A. (2004). Neuroplasticity: changes in grey matter induced by training. Nature *427*, 311–312.
10. Draganski, B., Gaser, C., Kempermann, G., Kuhn, H.G., Winkler, J., Büchel, C., and May, A. (2006). Temporal and spatial dynamics of brain structure changes during extensive learning. J. Neurosci. *26*, 6314–6317.
11. May, A., Hajak, G., Gänssbauer, S., Steffens, T., Langguth, B., Kleinjung, T., and Eichhammer, P. (2007). Structural brain alterations following 5 days of intervention: dynamic aspects of neuroplasticity. Cereb. Cortex *17*, 205–210.
12. Ilg, R., Wohlschläger, A.M., Gaser, C., Liebau, Y., Dauner, R., Wöller, A., Zimmer, C., Zihl, J., and Mühlau, M. (2008). Gray matter increase induced by practice correlates with task-specific activation: a combined functional and morphometric magnetic resonance imaging study. J. Neurosci. *28*, 4210–4215.
13. Boyke, J., Driemeyer, J., Gaser, C., Büchel, C., and May, A. (2008). Training-induced brain structure changes in the elderly. J. Neurosci. *28*, 7031–7035.
14. Ceccarelli, A., Rocca, M.A., Pagani, E., Falini, A., Comi, G., and Filippi, M. (2009). Cognitive learning is associated with gray matter changes in healthy human individuals: a tensor-based morphometry study. Neuroimage *48*, 585–589.
15. Hyde, K.L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A.C., and Schlaug, G. (2009). Musical training shapes structural brain development. J. Neurosci. *29*, 3019–3025.
16. Scholz, J., Klein, M.C., Behrens, T.E., and Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. Nat. Neurosci. *12*, 1370–1371.
17. Bezzola, L., Mérillat, S., Gaser, C., and Jäncke, L. (2011). Training-induced neural plasticity in golf novices. J. Neurosci. *31*, 12444–12448.
18. Erickson, K.I., Voss, M.W., Prakash, R.S., Basak, C., Szabo, A., Chaddock, L., Kim, J.S., Heo, S., Alves, H., White, S.M., et al. (2011). Exercise training increases size of hippocampus and improves memory. Proc. Natl. Acad. Sci. USA *108*, 3017–3022.

19. Maguire, E.A., Spiers, H.J., Good, C.D., Hartley, T., Frackowiak, R.S.J., and Burgess, N. (2003). Navigation expertise and the human hippocampus: a structural brain imaging analysis. Hippocampus *13*, 250–259.
20. Woollett, K., and Maguire, E.A. (2010). The effect of navigational expertise on wayfinding in new environments. J. Environ. Psychol. *30*, 565–573.
21. Woollett, K., Spiers, H.J., and Maguire, E.A. (2009). Talent in the taxi: a model system for exploring expertise. Philos. Trans. R. Soc. Lond. B Biol. Sci. *364*, 1407–1416.
22. Ashburner, J., and Friston, K.J. (2000). Voxel-based morphometry—the methods. Neuroimage *11*, 805–821.
23. Ashburner, J., and Friston, K.J. (2005). Unified segmentation. Neuroimage *26*, 839–851.
24. Kipps, C.M., Duggins, A.J., Mahant, N., Gomes, L., Ashburner, J., and McCusker, E.A. (2005). Progression of structural neuropathology in preclinical Huntington's disease: a tensor based morphometry study. J. Neurol. Neurosurg. Psychiatry *76*, 650–655.
25. O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. *34*, 171–175.
26. O'Keefe, J., and Nadel, L. (1978). The Hippocampus as a Cognitive Map (Oxford: Oxford University Press).
27. Burgess, N., Maguire, E.A., and O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. Neuron *35*, 625–641.
28. Moser, M.B., and Moser, E.I. (1998). Functional differentiation in the hippocampus. Hippocampus *8*, 608–619.
29. Fanselow, M.S., and Dong, H.W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? Neuron *65*, 7–19.
30. Lee, D.W., Miyasato, L.E., and Clayton, N.S. (1998). Neurobiological bases of spatial learning in the natural environment: neurogenesis and growth in the avian and mammalian hippocampus. Neuroreport *9*, R15–R27.
31. Sherry, D.F., and Hoshooley, J.S. (2010). Seasonal hippocampal plasticity in food-storing birds. Philos. Trans. R. Soc. Lond. B Biol. Sci. *365*, 933–943.
32. Shors, T.J. (2009). Saving new brain cells. Sci. Am. *300*, 46–52, 54.
33. Curlik, D.M., 2nd, and Shors, T.J. (2011). Learning increases the survival of newborn neurons provided that learning is difficult to achieve and successful. J. Cogn. Neurosci. *23*, 2159–2170.
34. Dayer, A.G., Ford, A.A., Cleaver, K.M., Yassaee, M., and Cameron, H.A. (2003). Short-term and long-term survival of new neurons in the rat dentate gyrus. J. Comp. Neurol. *460*, 563–572.
35. Doetsch, F., and Hen, R. (2005). Young and excitable: the function of new neurons in the adult mammalian brain. Curr. Opin. Neurobiol. *15*, 121–128.
36. Shors, T.J. (2008). From stem cells to grandmother cells: how neurogenesis relates to learning and memory. Cell Stem Cell *3*, 253–258.
37. Kolb, B., Gorny, G., Söderpalm, A.H., and Robinson, T.E. (2003). Environmental complexity has different effects on the structure of neurons in the prefrontal cortex versus the parietal cortex or nucleus accumbens. Synapse *48*, 149–153.
38. Toni, N., Teng, E.M., Bushong, E.A., Aimone, J.B., Zhao, C., Consiglio, A., van Praag, H., Martone, M.E., Ellisman, M.H., and Gage, F.H. (2007). Synapse formation on neurons born in the adult hippocampus. Nat. Neurosci. *10*, 727–734.
39. McEwen, B.S. (1999). Stress and hippocampal plasticity. Annu. Rev. Neurosci. *22*, 105–122.
40. Rakic, P. (2002). Neurogenesis in adult primate neocortex: an evaluation of the evidence. Nat. Rev. Neurosci. *3*, 65–71.
41. Anderson, J.C., Douglas, R.J., Martin, K.A., and Nelson, J.C. (1994). Map of the synapses formed with the dendrites of spiny stellate neurons of cat visual cortex. J. Comp. Neurol. *341*, 25–38.
42. Manganas, L.N., Zhang, X., Li, Y., Hazel, R.D., Smith, S.D., Wagshul, M.E., Henn, F., Benveniste, H., Djuric, P.M., Enikolopov, G., and Maletic-Savatic, M. (2007). Magnetic resonance spectroscopy identifies neural progenitor cells in the live human brain. Science *318*, 980–985.
43. Szeszko, P.R., Lipsky, R., Mentschel, C., Robinson, D., Gunduz-Bruce, H., Sevy, S., Ashtari, M., Napolitano, B., Bilder, R.M., Kane, J.M., et al. (2005). Brain-derived neurotrophic factor val66met polymorphism and volume of the hippocampal formation. Mol. Psychiatry *10*, 631–636.
44. Bueller, J.A., Aftab, M., Sen, S., Gomez-Hassan, D., Burmeister, M., and Zubieta, J.K. (2006). BDNF Val66Met allele is associated with reduced hippocampal volume in healthy subjects. Biol. Psychiatry *59*, 812–815.
45. Egan, M.F., Kojima, M., Callicott, J.H., Goldberg, T.E., Kolachana, B.S., Bertolino, A., Zaitsev, E., Gold, B., Goldman, D., Dean, M., et al. (2003). The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. Cell *112*, 257–269.
46. Hariri, A.R., Goldberg, T.E., Mattay, V.S., Kolachana, B.S., Callicott, J.H., Egan, M.F., and Weinberger, D.R. (2003). Brain-derived neurotrophic factor val66met polymorphism affects human memory-related hippocampal activity and predicts memory performance. J. Neurosci. *23*, 6690–6694.



Taylor & Francis
Taylor & Francis Group

Chronobiology International

The Journal of Biological and Medical Rhythm Research

ISSN: 0742-0528 (Print) 1525-6073 (Online) Journal homepage: https://www.tandfonline.com/loi/icbi20

Association between light exposure at night and manic symptoms in bipolar disorder: cross-sectional analysis of the APPLE cohort

Yuichi Esaki, Kenji Obayashi, Keigo Saeki, Kiyoshi Fujita, Nakao Iwata & Tsuyoshi Kitajima

To cite this article: Yuichi Esaki, Kenji Obayashi, Keigo Saeki, Kiyoshi Fujita, Nakao Iwata & Tsuyoshi Kitajima (2020): Association between light exposure at night and manic symptoms in bipolar disorder: cross-sectional analysis of the APPLE cohort, Chronobiology International, DOI: 10.1080/07420528.2020.1746799

To link to this article: https://doi.org/10.1080/07420528.2020.1746799

View supplementary material

Published online: 02 Apr 2020.

Submit your article to this journal

Article views: 7

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=icbi20

CHRONOBIOLOGY INTERNATIONAL
https://doi.org/10.1080/07420528.2020.1746799

Taylor & Francis
Taylor & Francis Group

Check for updates

Association between light exposure at night and manic symptoms in bipolar disorder: cross-sectional analysis of the APPLE cohort

Yuichi Esaki[a,b], Kenji Obayashi [c], Keigo Saeki[c], Kiyoshi Fujita[a,d], Nakao Iwata[b], and Tsuyoshi Kitajima[b]

[a]Department of Psychiatry, Okehazama Hospital, Aichi, Japan; [b]Department of Psychiatry, Fujita Health University School of Medicine, Aichi, Japan; [c]Department of Epidemiology, Nara Medical University School of Medicine, Nara, Japan; [d]Department of Psychiatry, The Neuroscience Research Center, Aichi, Japan

ABSTRACT

Previous studies have found that keeping the room dark at night was associated with a decrease in manic symptoms for patients with bipolar disorder (BD). However, the association between light at night of real-life conditions and manic symptoms is unclear. We investigated the association between bedroom light exposure at night and manic symptoms in BD patients. One-hundred and eighty-four outpatients with BD participated in this cross-sectional study. The average light intensity at night during sleep was evaluated using a portable photometer for seven consecutive nights. Manic symptoms were assessed using the Young Mania Rating Scale (YMRS), and scores ≥5 were treated as a "hypomanic state." The median (interquartile range) YMRS score was 2.0 (0–5.0), and 52 (28.2%) participants were in a hypomanic state. The prevalence of a hypomanic state was significantly higher in the participants with an average light intensity at night exposure of ≥3 lux than in those with <3 lux (36.7% versus 21.9%; $P = .02$). In multivariable logistic regression analysis adjusted for BD type, depressive symptoms, sleep duration, and daytime physical activity, the odds ratio (OR) for a hypomanic state was significantly higher for the participants with an average light intensity at night exposure of ≥3 lux than for those with <3 lux (OR: 2.15, 95% confidence interval: 1.09–4.22, $P = .02$). This association remained significant at the cutoff value of YMRS score ≥6 (OR: 2.51, 95% confidence interval: 1.15–5.46; $P = .02$). The findings of this study indicate bedroom light exposure at night is significantly associated with manic symptoms in BD patients. Although the results of this cross-sectional investigation do not necessarily imply causality, they may serve to inform beneficial nonpharmacological intervention and personalized treatment of BD patients.

ARTICLE HISTORY
Received 1 December 2019
Revised 7 March 2020
Accepted 8 March 2020

KEYWORDS
Bipolar disorder; manic symptom; light at night; dark; circadian rhythm

Introduction

Bipolar disorder (BD) is a severe chronic mental illness with a prevalence of 1–4%, which is characterized by episodes of depression and mania/hypomania (Merikangas et al. 2007). Manic or hypomanic episodes are states of elevated mood and increased activity or energy, and most patients complain of a reduced need for sleep during manic episodes (Harvey 2008). Patients with an acute manic episode often present with psychotic symptoms (Goodwin 2007), which increase the risk of a subsequent depressive episode (Salvadore et al. 2010). Symptoms during manic episodes cause impairment of social, economic, and professional functioning, leading to the possibility of hospitalization (Grande et al. 2016). Therefore, management of manic symptoms is of great clinical importance.

Light exposure, including seasonality and photoperiod, is closely associated with the pathophysiology of BD, such as altered mood symptoms, circadian rhythms, and sleep quality as well as risk suicide and hospitalization in emergency psychiatry (Aguglia et al. 2017, 2018; Geoffroy 2018; Aguglia et al. 2019; Bauer et al. 2019; Esaki et al. 2019a, b). For this reason, light regulation is widely used in the treatment of BD. Light therapy exposes patients to bright artificial light in the morning (Pail et al. 2011), and various studies have found that it is effective for bipolar depression (Sit et al. 2018; Tseng et al. 2016; Zhou et al. 2018). In contrast, darkness at night while sleeping is associated with reduction of manic symptoms in BD (Barbini et al. 2005). Findings from two case reports and one pilot study indicate that dark

CONTACT Yuichi Esaki esakiz@fujita-hu.ac.jp Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Aichi 4701192, Japan
Supplemental data for this article can be accessed publisher's website.
© 2020 Taylor & Francis Group, LLC

therapy (DT), which exposes patients to a dark room from 18:00 to 08:00 h, can exert striking beneficial effects on manic patients (Barbini et al. 2005; Wehr et al. 1998; Wirz-Justice et al. 1999). Additionally, a recent randomized placebo-controlled study demonstrated that virtual DT using blue-blocking glasses from 18:00 to 08:00 h significantly reduces mania scores relative to the scores attained with placebo glasses (Henriksen et al. 2016). Collectively, the results of these past investigations imply depression improves with morning/daytime light, whereas mania decreases with evening/nighttime darkness.

Although artificial light has great benefits for modern daily life, exposure to light at night (LAN) is associated with negative health effects, such as cancer, metabolic syndrome, mood disorders, and sleep disorders (Cho et al. 2015). It is likely that LAN exposure may affect manic symptoms in BD patients. However, to the best of our knowledge, no study has investigated the association between LAN exposure and manic symptoms in patients with BD. In this cross-sectional study, we evaluated the associations between bedroom LAN exposure and manic symptoms in patients with BD. We hypothesized that LAN exposure is associated with worsening of manic symptoms.

Materials and methods

Participants

Between August 2017 and January 2019, outpatients with BD participated in a study titled "Association between the Pathology of Bipolar Disorder and Light Exposure in Daily Life (APPLE) cohort study" from Okehazama Hospital, Fujita Mental Care Satellite Zengo, Fujita Mental Care Satellite Tokushige, and Fujita Health University. Eligible patients were 18–75 years old and had a diagnosis of BD I or II according to the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (American Psychiatric Association 2013). Primary analysis did not exclude pregnant patients or those who had other co-existing psychiatric diagnoses/conditions (e.g., personality disorders, substance abuse, and mental retardation), neurological disorders, and other diseases. Night-shift workers and those judged by a clinician to be at serious suicidal risk were excluded. A total of 370 participants were assessed for eligibility. Among them, 49 patients did not meet the inclusion criteria, and 131 patients refused to participate in the study. Therefore, a total of 190 patients were enrolled in this study. Five patients discontinued the study because of difficulty with the photometric measurement, and one patient could not be evaluated for mood symptoms. Finally, a total of 184 patients were analyzed in this study. The study procedure was approved by the ethics committee of Okehazama Hospital and was performed in accordance with the tenets of the Declaration of Helsinki and the ethical standards of the journal (Portaluppi et al. 2010). All participating patients provided written informed consent, and the study was registered at UMIN-CTR (identifier: UMIN000028239).

Study procedure

We evaluated the demographic and clinical characteristics of the participants at the clinic, and then instructed them to perform the following procedures and measurements for 7 consecutive days: (1) record bedroom light from bedtime to rising time using a portable photometer (LX-28SD; Sato Shouji Inc., Kanagawa, Japan) and (2) record bedtime and rising time in a sleep diary. Bedtime was defined as the point in time when the participant was in bed and intended to go to sleep and did not start if the participant was in bed reading books or watching TV. Rising time was defined as the point in time when the participant actually got out of bed.

LAN exposure assessment

LAN exposure in the bedroom was recorded at 1-min intervals using the photometer. We instructed the participant to position the light sensor of the photometer near their head at eye level and to start the recording at bedtime and end it at final rising time. We defined the average light intensity at night over the entire period from bedtime to rising time as the average LAN exposure. A previous experimental study reported that a LAN exposure ≥3 lux affected the circadian rhythm of healthy adults (Zeitzer et al. 2000). Additionally, another experimental study has

shown that LAN of ≥5 lux during sleep might affect the sleep architecture of healthy young men (Cho et al. 2016). Thus, we used the value of 3 lux as a cutoff threshold in this study. The procedure for measuring LAN exposure has been used in previous studies of the general population as well as patients with BD (Esaki et al. 2019b; Obayashi et al. 2012).

Manic symptoms' assessment

Manic symptoms were evaluated at the start of the experimental period using the Young Mania Rating Scale (YMRS) (Young et al. 1978). The YMRS is an 11-item diagnostic questionnaire used to measure the severity of manic symptoms in patients with BD; the overall score ranges from 0 to 60. A higher YMRS score indicates more severe mania. The International Society for Bipolar Disorders (ISBD) Task Force recommends a YMRS score of <5 as symptomatic remission of bipolar mania (Tohen et al. 2009). Therefore, as used herein, a YMRS score ≥5 points is indicative of a "hypomanic state." Additionally, we also used different cut off points (YMRS score ≥6 and ≥13) because a definition of remission in BD varied and were inconsistent.

Other assessment

Depressive symptoms were evaluated at the start of the experimental period using the Montgomery–Åsberg Depression Rating Scale (MADRS) (Montgomery and Asberg 1979). Sleep duration, daytime physical activity, and daytime light exposure were evaluated using actigraphy (Actiwatch Spectrum Plus; Respironics Inc., Pittsburgh, PA, USA). The participants were instructed to wear an actigraph on the wrist of their non-dominant arm 24 h/day, except while bathing and during aquatic activities, for 7 consecutive days. We used a default threshold as wake (≥40 counts per min) and sleep periods (<40 counts per min). This threshold setting has been validated to be as accurate as polysomnography in measuring sleep in patients with BD (Kaplan et al. 2012). Sleep duration was the total time from sleep start to end. Daytime physical activity was defined as the average activity count per minute from rising time to bedtime. Average daytime light was defined as the average daytime light intensity from rising time to bedtime. The information on psychiatric medications, including mood stabilizers (lithium, lamotrigine, valproate, and carbamazepine), antipsychotics, antidepressants, and hypnotics were obtained from the clinical records. Data of the day length from sunrise to sunset in Aichi (latitude, 35° N) on the first measurement day of each subject were collected from the web page of the National Astronomical Observatory of Japan. Sociodemographics, including married status and employment status, were obtained from the questionnaire. Clinical characteristics, including type of BD, age at onset of BD, duration of illness, and family history of psychiatric disorders, were obtained from the clinical records or the participants' doctors.

Statistical analyses

Normally distributed variables are presented as the mean and standard deviation (SD), those with an asymmetrical distribution as median and interquartile range (IQR), and categorical variables as number and percentage. The night-to-night correlations of LAN exposure during the 7 days study span were evaluated using the Spearman's rank correlation coefficient (*rs*). In comparisons between the hypomanic state and non-hypomanic state groups, unpaired t-tests, Mann–Whitney U test, Fisher's exact test, and chi-square test were used to compare their demographic characteristics (age, sex, marital status, and employment), clinical characteristics (type of BD, age at onset of BD, duration of illness, psychiatric comorbidity, family history, MADRS and YMRS scores, bedtime, rising time, sleep duration, daytime physical activity, average daytime light, and day length), and psychiatric medications (mood stabilizers, antipsychotics, antidepressants, and hypnotics).

Associations of average LAN (≥3 versus <3 lux) with manic symptoms (median YMRS scores and prevalence of hypomanic state [YMRS score ≥5], YMRS score ≥6 and ≥13) were evaluated using the Mann–Whitney U test and chi-square test, respectively. Odds ratios (ORs) for a hypomanic state and YMRS score ≥6 were evaluated using the univariable and multivariable logistic regression model, respectively. The MADRS score was not normally distributed, even after log transformation. It was, therefore, treated as a categorical variable in the logistic

regression analyses as being either above or below a cutoff value of 8 points (the cutoff value for depression recommended by the ISBD Task Force) (Tohen et al. 2009). In the multivariable model, the ORs were simultaneously adjusted for variables associated with mania chosen based on previous studies, including type of BD (I or II) (Judd et al. 2003, 2002), MADRS score (≥8 or <8) (Judd et al. 2003, 2002), sleep duration (in min) (Salvatore et al. 2008), and daytime physical activity (in count/min) (Krane-Gartiser et al. 2014). Statistical analyses were performed using SPSS version 25.0 for Windows (IBM Corp., Armonk, NY). A two-sided p value < .05 was considered as indicative of statistical significance.

Results

Of the total 184 participants, 103 (56%) were female, and 81 (44%) were male, with a mean (SD) age of 45.1 (13.1) years. The median (IQR) of the YMRS scores was 2.0 (0–5.0), and 52 (28.2%) participants were classified as being in a hypomanic state during the 7-day study period. Among these 52 participants, 19 (36.6%) were in a hypomanic state (YMRS score ≥5 and MADRS score <8), and 33 (63.4%) were in a mixed state, including a depressive state (YMRS score ≥5 and MADRS score ≥8). The median (IQR) of the average LAN exposure was 2.0 (0.3–9.5) lux; 79 (42.9%) participants had an average LAN exposure of ≥3 lux, and 105 (57.1%) had exposure of <3 lux. The night-to-night correlations of the LAN exposure throughout the 7 days study span were moderate (*rs* range: 0.55–0.72).

As shown in Table 1, the median MADRS score was significantly higher in the hypomanic state group than in the non-hypomanic state group (12.0 versus 8.0; P = .03). Sleep duration was significantly shorter in the hypomanic state group than in the non-hypomanic state group (413.4 versus 450.6 min; P = .01).

The prevalence of the hypomanic state (YMRS score ≥ 5) was significantly higher in the participants with average LAN ≥ 3 lux than in those with <3 lux (36.7% versus 21.9%; P = .02, Figure 1). This association remained significant at different cutoff values of the YMRS score (YMRS score ≥6: 27.3% versus 14.3%; P = .02, YMRS score ≥13: 8.9% versus 1.0%; P = .009). Consistently, the median (IQR) YMRS score was 3.0 (1.0–6.0) in the participants with an average LAN exposure of ≥3 lux and was 2.0 (0–4.0) in those with <3 lux (P < .01). Additional analysis excluding participants with psychiatric comorbidity (n = 48) suggested that these associations remained significant (hypomanic state: 34.5% versus 17.9%; P = .02; YMRS score ≥ 6: 29.3% versus 12.8%; P = .01; YMRS score ≥13: 10.3% versus 1.3%; P = .01; median YMRS score: P = .01).

In univariable logistic regression analysis, crude ORs for a hypomanic state were significantly associated with the average LAN exposure and sleep duration (LAN: OR, 2.068; 95% CI: 1.079–3.963, P = .029; sleep duration: OR, 0.996; 95% CI: 0.992–0.999, P = .021; Table 2). In multivariable logistic regression analysis after adjusting for type of BD, MADRS score, sleep duration, and daytime physical activity, increased average LAN exposure was associated with higher adjusted OR for a hypomanic state (OR, 2.152, 95% CI: 1.095–4.228; P = .026, Table 2). Consistent with this, the crude and adjusted OR for YMRS score ≥6 were significantly associated with average LAN exposure (crude OR, 2.316, 95% CI: 1.110–4.831; P = .025; adjusted OR, 2.517, 95% CI: 1.158–5.468; P = .020, Supplemental Table 1).

Discussion

To our knowledge, this is the first study that has investigated the associations between LAN exposure in the real-life situation and manic symptoms in BD. We demonstrated that average LAN intensity was significantly associated with manic symptoms in patients with BD. This association was independent of potentially confounding factors, including type of BD, depressive symptoms, sleep duration, and daytime physical activity.

The study results are consistent with those of previous studies on the effectiveness of DT in bipolar mania. One previous study reported that adjunctive DT achieved a 50% reduction in the YMRS scores in 3 days for 16 patients with an acute manic episode (Barbini et al. 2005). Two case reports also showed that DT rapidly improved mood swing in rapid cycling patients (Wehr et al. 1998; Wirz-Justice et al. 1999). Furthermore, a recent randomized controlled trial involving 32 patients with bipolar mania demonstrated that

Table 1. Demographic and clinical characteristics between the hypomanic state and non-hypomanic state groups.

Variables	Hypomanic state (n = 52)	Non-hypomanic state (n = 132)	P
Demographic characteristics			
Age, years, mean (SD)	45.5 (12.0)	44.9 (13.6)	0.76
Sex, female, n (%)	32 (61.5)	71 (53.8)	0.34
Married, n (%)	29 (55.8)	68 (51.5)	0.60
Employed, n (%)	16 (31.4)	57 (43.2)	0.14
Clinical characteristics			
Type of BD, BD I, n (%)	21 (40.4)	45 (34.1)	0.42
Age at onset of BD, years, mean (SD)	31.2 (10.5)	32.9 (12.8)	0.39
Duration of illness, years, mean (SD)	14.3 (9.4)	12.0 (9.3)	0.14
Psychiatric comorbidity, n (%)	18 (34.6)	30 (22.7)	0.09
Family history of psychiatric disorders, n (%)	10 (19.2)	21 (15.9)	0.58
MADRS score, points, median (IQR)	12.0 (6.0–16.7)	8.0 (2.0–13.0)	0.03
YMRS score, points, median (IQR)	7.0 (5.0–9.0)	1.0 (0–2.0)	<0.01
Bedtime, clock time, mean (SD)	23:28 (1:58)	23:07 (1:35)	0.22
Rising time, clock time, mean (SD)	7:02 (2:03)	7:18 (1:47)	0.38
Sleep duration, min, mean (SD)	413.4 (84.0)	450.6 (100.0)	0.01
Daytime physical activity, counts/min, mean (SD)	188.1 (68.6)	171.1 (69.4)	0.15
Average daytime light, lux, median (IQR)	244.3 (175.2–315.5)	217.0 (150.3–308.4)	0.28
Day length, min, median (IQR)	729 (686–816)	759 (670–836)	0.42
Medications			
Lithium, n (%)	22 (42.3)	52 (39.4)	0.71
Lamotrigine, n (%)	19 (36.5)	41 (31.1)	0.47
Valproate, n (%)	11 (21.2)	34 (25.8)	0.51
Carbamazepine, n (%)	3 (5.8)	1 (0.8)	0.06
Atypical antipsychotics, n (%)	30 (57.7)	71 (53.8)	0.63
Antidepressants, n (%)	19 (36.5)	45 (34.1)	0.75
Hypnotics, n (%)	34 (65.4)	82 (62.1)	0.68

Data are expressed as mean (standard deviation, SD), median (interquartile range, IQR), or number (proportion). MADRS, Montgomery–Åsberg Depression Rating Scale; YMRS, Young Mania Rating Scale

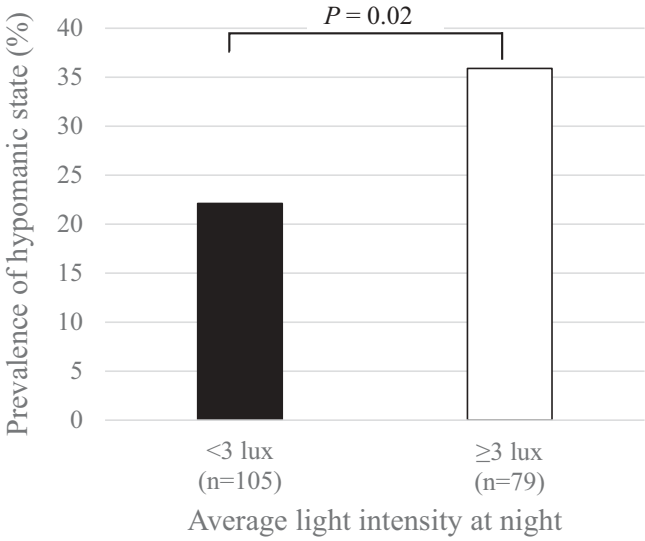


Figure 1. Associations of average light intensity (≥3 versus <3 lux) at night with prevalence of hypomanic state. P value is for the analysis using the chi-square test.

blue-blocking glasses therapy, which is virtual DT, relative to placebo glasses significantly improved YMRS score (Cohen's d = 1.86) (Henriksen et al. 2016). Our results indicated participants exposed to an average LAN ≥3 lux had an OR for manic symptom approximately two times higher than that for those exposed to an average LAN <3 lux. Although more than half the participants in the hypomanic state group had depressive symptoms (mixed state), the multivariable analysis demonstrated that the association between LAN exposure and manic symptoms were independent of depressive symptoms.

We also observed that shorter sleep duration was associated with increased mania, independently of average LAN, age, type of BD, and daytime physical activity. In a previous review study, 69–99% of the patients during the manic state complained of a reduced need for sleep in self-reported data, and polysomnography confirmed the decreased time spent asleep and shortened rapid eye movement sleep latency in the manic state (Harvey 2008). Additionally, another study reported that 32 patients with manic/mixed state experienced shorter total sleep time measured by actigraphy than that of 36 patients in a recovery

Table 2. Logistic regression analysis for the associations of each variable with odds ratios (ORs) for a hypomanic state.

Variables	Crude OR	95% CI		*P*	Adjusted OR	95% CI		*P*
Average LAN (≥3 versus <3 lux)	2.068	1.079	3.963	0.029	2.152	1.095	4.228	0.026
Type of BD (I versus II)	1.30	0.67	2.53	0.42	1.38	0.68	2.78	0.36
MADRS score (≥8 or <8 points)	1.63	0.84	3.16	0.14	1.71	0.84	3.50	0.13
Sleep duration (per min)	0.996	0.992	0.999	0.021	0.996	0.992	1.000	0.039
Daytime physical activity (per counts/min)	1.00	0.99	1.00	0.15	1.004	0.99	1.00	0.36

OR, odds ratio; CI, confidence interval; LAN, light at night. Adjusted for all covariates shown.

state (mean; 7.03 versus 8.54 h) (Salvatore et al. 2008). Our results showed that the participants in a hypomanic state had shorter sleep duration than that of the participants in a non-hypomanic state (mean; 413 versus 450 min). Therefore, our results are consistent with those of the previous study.

Our findings indicate that maintaining darkness during sleep could be a novel clinical therapeutic option for BD. Although pharmacological and nonpharmacological treatments for BD are progressing, one-third to one-half of patients with BD experience recurrence within 1 year (Gitlin et al. 1995; Keller et al. 1993; Perlis et al. 2006). A new treatment option is needed because, although DT is effective in bipolar mania, it is difficult for manic patients to maintain a dark room from 18:00 to 08:00 h (Wirz-Justice and Terman 2016). We consider that it may be possible for manic patients to maintain bedroom darkness only during sleep. Therefore, maintaining darkness during sleep is expected to have a beneficial or even curative effect on manic patients and a preventive effect on relapse of manic episodes. Further research using a longitudinal controlled study design is needed to determine the association between LAN exposure and manic symptoms.

The mechanism underlying the association between LAN exposure and manic symptoms in BD remains unknown, but it may be related to circadian rhythm dysfunction and sleep disturbance. Light exposure has a primary role in the regulation of circadian rhythms in humans (Dumont and Beaulieu 2007), and LAN exposure causes disruption of the biological clock and suppression of melatonin secretion (Lunn et al. 2017; Reiter et al. 2007; Smolensky et al. 2016, 2015). Furthermore, BD has been suggested to associate with the abnormality of circadian rhythm, including evening preference, light hypersensitivity, and the abnormal of melatonin and cortisol secretions (Bellivier et al. 2015; Geoffroy 2018). A previous longitudinal study demonstrated that BD patients with circadian rhythm sleep disorder (CRSD) had higher relapse of depressive, manic, or hypomanic episodes than those without CRSD (Takaesu et al. 2018). Therefore, circadian rhythm dysfunction is one of the major factors of the association between LAN exposure and mania. Another possible mechanism is sleep disturbance. Although LAN exposure is linked to variable negative health effects, sleep disturbance is one of the most obvious negative effects of LAN exposure (Cho et al. 2015). The association between LAN exposure and sleep disturbance has been reported not only from an experimental study but also from a study involving subjects with real-life situations (Cho et al. 2016; Obayashi et al. 2014). Sleep disturbance, particularly sleep loss, is associated with the risk for manic/hypomanic episodes (Lewis et al. 2018, 2017). Similarly, we observed a high OR for manic symptomatology in the participants with shorter sleep duration (Table 2). Therefore, it is likely that sleep disturbance may be involved in the association between LAN exposure and manic symptoms, although this association was independent of sleep duration measured using actigraphy. Various methods of evaluating sleep have been performed, including sleep diary, questionnaire, and polysomnography (PSG); in particular, PSG can evaluate not only sleep continuity but also its architecture. Thus, further studies using PSG are needed to clarify whether sleep mediates between LAN exposure and manic symptoms in BD.

This study had several limitations. First, this was a cross-sectional study; therefore, the results do not necessarily imply that LAN exposure worsens manic symptoms in patients with BD. Similarly, the causal relationship of LAN exposure and sleep duration is also unclear. Further longitudinal investigation is necessary to clarify these associations. Second, the

study participants were not randomly selected. The results may therefore have been affected by selection bias. Third, the YMRS score of participants in our study was low. In previous conducted DT and virtual DT studies, mean (SD) YMRS scores at the start of intervention were, respectively, 31.62 (10.92) and 23.4 (8.0) (Barbini et al. 2005; Henriksen et al. 2016). On the other hand, the median (IQR) YMRS score in the hypomanic state group of our study was 7.0 (5.0–9.0). It was difficult to recruit outpatients with acute manic symptoms in our study, because they are usually hospitalized until their condition improves. Therefore, this factor may have led to the underestimation of the association between LAN exposure and manic symptoms. Additionally, although we defined YMRS score ≥5 as a "hypomanic state" in accord with the recommendations of the ISBD (Tohen et al. 2009), the definition and threshold YMRS of previous studies varied and were inconsistent because of insufficient consensus concerning the definition of remission in BD. Therefore, we explored the impact of different cutoff YMRS thresholds on the findings. Our results indicated that different cutoff value of the YMRS score (YMRS score ≥6 versus <6) were also significantly associated with average LAN exposure. Fourth, our primary analysis did not exclude patients with other psychiatric disorders, neurological disorders, and pregnancy. However, the results of additional analysis excluding participants with psychiatric comorbidity were consistent with those in the primary analysis. Fifth, LAN exposure was assessed only by measuring light intensity, not wavelength, by a photometer positioned at eye level. Previous experimental studies have reported that the circadian time structure is not affected by longer light wavelengths (red, orange, or yellow–green), but only by shorter wavelengths (blue and blue–green) (Brainard et al. 2001; Thapan et al. 2001). Another previous study demonstrated that wearing glasses that block blue wavelengths was effective for acute manic symptoms (Henriksen et al. 2016). Therefore, future studies measuring wavelengths could reveal a more appropriate association between LAN exposure and mania. In addition, we had no data on light exposure if the participant left the bedroom at night, such as to visit the bathroom. Moreover, although light data sensed by the wrist actigraphy were recorded 24 h/day for 7 consecutive days, such sleep-time data were excluded from the analysis, because they were considered to contain several artifacts caused by clothing or bedding covering the light sensor of the actigraph. Therefore, our analysis relied only on LAN intensity measured by the photometer in the bedroom from bedtime to rising time. A previous study demonstrated that patients in a manic state had a higher proportion of nocturnal activity measured by actigraphy than that of patients in an euthymic state (Salvatore et al. 2008). Therefore, the participants in the hypomanic state group may have been exposed to more LAN than reflected by the measured results. Moreover, the source of the light could not be distinguished, such as bedroom light or outdoor light entering the bedroom. Furthermore, LAN exposure was only measured for 7 days, so the average LAN intensity over the 7 days may not be the accurate representative value for that period. However, the night-to-night correlations of LAN exposure over 7 days were moderately high, which suggests that the average LAN intensity may be acceptable as the representative value for 7 days. Finally, although we demonstrated that the association between LAN exposure and manic symptoms was independent of several potential confounding factors, we did not investigate other potential confounding factors such as features of BD (e.g., predominant polarity, bipolar cycle, and number of affective episodes), psychological factors, environmental factors (e.g., weather and noise at night), and pharmacological treatment. In addition, the present study lacked a control group.

In conclusion, the study results demonstrated that LAN exposure was significantly associated with manic symptoms in BD patients. This finding may have potential benefit for non-pharmacological intervention and personalized treatment of BD in the future. Further investigation is necessary to confirm the association between LAN exposure and bipolar mania.

Acknowledgements

We are grateful to the patients who participated in this study. We also thank Soji Tsuboi and Miyuki Yamamoto for their valuable support during this research.

Declaration of Interest statement

The authors report no conflicts of interest related to this research. Dr Obayashi and Dr Saeki has received a research grant from

YKK AP Inc.; Ushio Inc.; Tokyo Electric Power Company; EnviroLife Research Institute Co., Ltd.; Sekisui Chemical Co., Ltd; LIXIL Corp.; and KYOCERA Corp. Dr Fujita has received speaker's honoraria from Dainippon Sumitomo, Eli Lilly, GlaxoSmithKline, Janssen, Yoshitomi, Otsuka, Meiji, Shionogi, Novartis, and Kracie. Dr Iwata has received speaker's honoraria from Astellas, Dainippon Sumitomo, Eli Lilly, GlaxoSmithKline, Janssen, Yoshitomi, Otsuka, Meiji, Shionogi, Novartis, and Pfizer and has had research grants from GlaxoSmithKline, Meiji, Otsuka, Mitsubishi Tanabe, Dainippon Sumitomo, Daiichisankyo, and Eisai. Dr Kitajima has received speaker's honoraria from Eisai, Mitsubishi Tanabe, Otsuka, Takeda, Eli Lilly, MSD, Meiji, Yoshitomi, Fukuda, Dainippon Sumitomo, Shionogi, and Novo Nordisk, and has received a research grant from Eisai, MSD and Takeda.

Funding

This work was supported by the Japan Foundation for Neuroscience and Mental Health; The Neuroscience Research Center; Scientific Research from JSPS KAKENHI [18K15529].

ORCID

Kenji Obayashi http://orcid.org/0000-0001-6346-9082

References

Aguglia A, Borsotti A, Cuniberti F, Serafini G, Amore M, Maina G. 2017. The influence of sunlight exposure on hospitalization in emergency psychiatry. Chronobiol Int. 34(10):1413–1422. doi:10.1080/07420528.2017.1374286.

Aguglia A, Borsotti A, Maina G. 2018. Bipolar disorders: is there an influence of seasonality or photoperiod? Braz J Psychiatry. 40(1):6–11. doi:10.1590/1516-4446-2016-2144.

Aguglia A, Serafini G, Solano P, Giacomini G, Conigliaro C, Salvi V, Mencacci C, Romano M, Aguglia E, Amore M. 2019. The role of seasonality and photoperiod on the lethality of suicide attempts: A case-control study. J Affect Disord. 246:895–901. doi:10.1016/j.jad.2018.12.094.

American Psychiatric Association. 2013. Diagnostic and statistical manual of mental disorders. 5th ed. American Psychiatric Publishing.

Barbini B, Benedetti F, Colombo C, Dotoli D, Bernasconi A, Cigala-Fulgosi M, Florita M, Smeraldi E. 2005. Dark therapy for mania: a pilot study. Bipolar Disord. 7(1):98–101. doi:10.1111/j.1399-5618.2004.00166.x.

Bauer M, Glenn T, Alda M, Andreassen OA, Angelopoulos E, Ardau R, Ayhan Y, Baethge C, Bauer R, Baune BT, et al. 2019. Association between solar insolation and a history of suicide attempts in bipolar I disorder. J Psychiatr Res. 113:1–9. doi:10.1016/j.jpsychires.2019.03.001.

Bellivier F, Geoffroy PA, Etain B, Scott J. 2015. Sleep- and circadian rhythm-associated pathways as therapeutic targets in bipolar disorder. Expert Opin Ther Targets. 19(6):747–763. doi:10.1517/14728222.2015.1018822.

Brainard GC, Hanifin JP, Greeson JM, Byrne B, Glickman G, Gerner E, Rollag MD. 2001. Action spectrum for melatonin regulation in humans: evidence for a novel circadian photoreceptor. J Neurosci. 21(16):6405–6412. doi:10.1523/JNEUROSCI.21-16-06405.2001.

Cho CH, Lee HJ, Yoon HK, Kang SG, Bok KN, Jung KY, Kim L, Lee EI. 2016. Exposure to dim artificial light at night increases REM sleep and awakenings in humans. Chronobiol Int. 33(1):117–123. doi:10.3109/07420528.2015.1108980.

Cho Y, Ryu SH, Lee BR, Kim KH, Lee E, Choi J. 2015. Effects of artificial light at night on human health: A literature review of observational and experimental studies applied to exposure assessment. Chronobiol Int. 32(9):1294–1310. doi:10.3109/07420528.2015.1073158.

Dumont M, Beaulieu C. 2007. Light exposure in the natural environment: relevance to mood and sleep disorders. Sleep Med. 8(6):557–565. doi:10.1016/j.sleep.2006.11.008.

Esaki Y, Kitajima T, Obayashi K, Saeki K, Fujita K, Iwata N. 2019a. Daytime light exposure in daily life and depressive symptoms in bipolar disorder: A cross-sectional analysis in the APPLE cohort. J Psychiatr Res. 116:151–156. doi:10.1016/j.jpsychires.2019.06.010.

Esaki Y, Kitajima T, Obayashi K, Saeki K, Fujita K, Iwata N. 2019b. Light exposure at night and sleep quality in bipolar disorder: the APPLE cohort study. J Affect Disord. 257:314–320. doi:10.1016/j.jad.2019.07.031.

Geoffroy PA. 2018. Clock genes and light signaling alterations in bipolar disorder: when the biological clock is off. Biol Psychiatry. 84(11):775–777. doi:10.1016/j.biopsych.2018.09.006.

Gitlin MJ, Swendsen J, Heller TL, Hammen C. 1995. Relapse and impairment in bipolar disorder. Am J Psychiatry. 152(11):1635–1640. doi:10.1176/ajp.152.11.1635.

Goodwin FJK. 2007. Manic-depressive illness: bipolar disorders and recurrent depression. 2nd ed. New York: Oxford University Press.

Grande I, Berk M, Birmaher B, Vieta E. 2016. Bipolar disorder. Lancet. 387(10027):1561–1572. doi:10.1016/S0140-6736(15)00241-X.

Harvey AG. 2008. Sleep and circadian rhythms in bipolar disorder: seeking synchrony, harmony, and regulation. Am J Psychiatry. 165(7):820–829. doi:10.1176/appi.ajp.2008.08010098.

Henriksen TE, Skrede S, Fasmer OB, Schoeyen H, Leskauskaite I, Bjorke-Bertheussen J, Assmus J, Hamre B, Gronli J, Lund A. 2016. Blue-blocking glasses as additive treatment for mania: a randomized placebo-controlled trial. Bipolar Disord. 18(3):221–232. doi:10.1111/bdi.12390.

Judd LL, Akiskal HS, Schettler PJ, Coryell W, Endicott J, Maser JD, Solomon DA, Leon AC, Keller MB. 2003. A prospective investigation of the natural history of the long-term weekly symptomatic status of bipolar II

disorder. Arch Gen Psychiatry. 60(3):261–269. doi:10.1001/archpsyc.60.3.261.

Judd LL, Akiskal HS, Schettler PJ, Endicott J, Maser J, Solomon DA, Leon AC, Rice JA, Keller MB. 2002. The long-term natural history of the weekly symptomatic status of bipolar I disorder. Arch Gen Psychiatry. 59(6):530–537. doi:10.1001/archpsyc.59.6.530.

Kaplan KA, Talbot LS, Gruber J, Harvey AG. 2012. Evaluating sleep in bipolar disorder: comparison between actigraphy, polysomnography, and sleep diary. Bipolar Disord. 14(8):870–879. doi:10.1111/bdi.12021.

Keller MB, Lavori PW, Coryell W, Endicott J, Mueller TI. 1993. Bipolar I: a five-year prospective follow-up. J Nerv Ment Dis. 181(4):238–245. doi:10.1097/00005053-199304000-00005

Krane-Gartiser K, Henriksen TE, Morken G, Vaaler A, Fasmer OB. 2014. Actigraphic assessment of motor activity in acutely admitted inpatients with bipolar disorder. PLoS One. 9(2):e89574. doi:10.1371/journal.pone.0089574.

Lewis KJS, Di Florio A, Forty L, Gordon-Smith K, Perry A, Craddock N, Jones L, Jones I. 2018. Mania triggered by sleep loss and risk of postpartum psychosis in women with bipolar disorder. J Affect Disord. 225:624–629. doi:10.1016/j.jad.2017.08.054.

Lewis KS, Gordon-Smith K, Forty L, Di Florio A, Craddock N, Jones L, Jones I. 2017. Sleep loss as a trigger of mood episodes in bipolar disorder: individual differences based on diagnostic subtype and gender. Br J Psychiatry. 211(3):169–174. doi:10.1192/bjp.bp.117.202259.

Lunn RM, Blask DE, Coogan AN, Figueiro MG, Gorman MR, Hall JE, Hansen J, Nelson RJ, Panda S, Smolensky MH, et al. 2017. Health consequences of electric lighting practices in the modern world: A report on the National Toxicology Program's workshop on shift work at night, artificial light at night, and circadian disruption. Sci Total Environ. 607–608:1073–1084. doi:10.1016/j.scitotenv.2017.07.056.

Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RM, Petukhova M, Kessler RC. 2007. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. Arch Gen Psychiatry. 64(5):543–552. doi:10.1001/archpsyc.64.5.543.

Montgomery SA, Asberg M. 1979. A new depression scale designed to be sensitive to change. Br J Psychiatry. 134:382–389. doi:10.1192/bjp.134.4.382.

Obayashi K, Saeki K, Iwamoto J, Okamoto N, Tomioka K, Nezu S, Ikada Y, Kurumatani N. 2012. Positive effect of daylight exposure on nocturnal urinary melatonin excretion in the elderly: a cross-sectional analysis of the HEIJO-KYO study. J Clin Endocrinol Metab. 97(11):4166–4173. doi:10.1210/jc.2012-1873.

Obayashi K, Saeki K, Kurumatani N. 2014. Association between light exposure at night and insomnia in the general elderly population: the HEIJO-KYO cohort. Chronobiol Int. 31(9):976–982. doi:10.3109/07420528.2014.937491.

Pail G, Huf W, Pjrek E, Winkler D, Willeit M, Praschak-Rieder N, Kasper S. 2011. Bright-light therapy in the treatment of mood disorders. Neuropsychobiology. 64(3):152–162. doi:10.1159/000328950.

Perlis RH, Ostacher MJ, Patel JK, Marangell LB, Zhang H, Wisniewski SR, Ketter TA, Miklowitz DJ, Otto MW, Gyulai L, et al. 2006. Predictors of recurrence in bipolar disorder: primary outcomes from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). Am J Psychiatry. 163(2):217–224. doi:10.1176/appi.ajp.163.2.217

Portaluppi F, Smolensky MH, Touitou Y. 2010. Ethics and methods for biological rhythm research on animals and human beings. Chronobiol Int. 27(9–10):1911–1929. doi:10.3109/07420528.2010.516381.

Reiter RJ, Tan DX, Korkmaz A, Erren TC, Piekarski C, Tamura H, Manchester LC. 2007. Light at night, chronodisruption, melatonin suppression, and cancer risk: a review. Crit Rev Oncog. 13(4):303–328. doi:10.1615/CritRevOncog.v13.i4.

Salvadore G, Quiroz JA, Machado-Vieira R, Henter ID, Manji HK, Zarate CA Jr. 2010. The neurobiology of the switch process in bipolar disorder: a review. J Clin Psychiatry. 71(11):1488–1501. doi:10.4088/JCP.09r05259gre.

Salvatore P, Ghidini S, Zita G, De Panfilis C, Lambertino S, Maggini C, Baldessarini RJ. 2008. Circadian activity rhythm abnormalities in ill and recovered bipolar I disorder patients. Bipolar Disord. 10(2):256–265. doi:10.1111/j.1399-5618.2007.00505.x.

Sit DK, McGowan J, Wiltrout C, Diler RS, Dills JJ, Luther J, Yang A, Ciolino JD, Seltman H, Wisniewski SR, et al. 2018. Adjunctive bright light therapy for bipolar depression: a randomized Double-Blind placebo-controlled trial. Am J Psychiatry. 175(2):131–139. doi:10.1176/appi.ajp.2017.16101200

Smolensky MH, Hermida RC, Reinberg A, Sackett-Lundeen L, Portaluppi F. 2016. Circadian disruption: new clinical perspective of disease pathology and basis for chronotherapeutic intervention. Chronobiol Int. 33(8):1101–1119. doi:10.1080/07420528.2016.1184678.

Smolensky MH, Sackett-Lundeen LL, Portaluppi F. 2015. Nocturnal light pollution and underexposure to daytime sunlight: complementary mechanisms of circadian disruption and related diseases. Chronobiol Int. 32(8):1029–1048. doi:10.3109/07420528.2015.1072002.

Takaesu Y, Inoue Y, Ono K, Murakoshi A, Futenma K, Komada Y, Inoue T. 2018. Circadian rhythm sleep-wake disorders predict shorter time to relapse of mood episodes in euthymic patients with bipolar disorder: a prospective 48-week study. J Clin Psychiatry. 79:1. doi:10.4088/JCP.17m11565.

Thapan K, Arendt J, Skene DJ. 2001. An action spectrum for melatonin suppression: evidence for a novel non-rod, non-cone photoreceptor system in humans. J Physiol. 535(Pt 1):261–267. doi:10.1111/tjp.2001.535.issue-1.

Tohen M, Frank E, Bowden CL, Colom F, Ghaemi SN, Yatham LN, Malhi GS, Calabrese JR, Nolen WA, Vieta E, et al. 2009. The International Society for Bipolar Disorders

(ISBD) Task Force report on the nomenclature of course and outcome in bipolar disorders. Bipolar Disord. 11 (5):453–473. doi:10.1111/j.1399-5618.2009.00726.x

Tseng PT, Chen YW, Tu KY, Chung W, Wang HY, Wu CK, Lin PY. 2016. Light therapy in the treatment of patients with bipolar depression: A meta-analytic study. Eur Neuropsychopharmacol. 26(6):1037–1047. doi:10.1016/j.euroneuro.2016.03.001.

Wehr TA, Turner EH, Shimada JM, Lowe CH, Barker C, Leibenluft E. 1998. Treatment of rapidly cycling bipolar patient by using extended bed rest and darkness to stabilize the timing and duration of sleep. Biol Psychiatry. 43(11):822–828. doi:10.1016/S0006-3223(97)00542-8.

Wirz-Justice A, Quinto C, Cajochen C, Werth E, Hock C. 1999. A rapid-cycling bipolar patient treated with long nights, bedrest, and light. Biol Psychiatry. 45 (8):1075–1077. doi:10.1016/S0006-3223(98)00289-3.

Wirz-Justice A, Terman M. 2016. Commentary on "Blue-blocking glasses as additive treatment for mania: a randomized placebo-controlled trial". Bipolar Disord. 18(4):383–384. doi:10.1111/bdi.12392.

Young RC, Biggs JT, Ziegler VE, Meyer DA. 1978. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry. 133:429–435. doi:10.1192/bjp.133.5.429.

Zeitzer JM, Dijk DJ, Kronauer R, Brown E, Czeisler C. 2000. Sensitivity of the human circadian pacemaker to nocturnal light: melatonin phase resetting and suppression. J Physiol. 526(Pt 3):695–702. doi:10.1111/j.1469-7793.2000.00695.x.

Zhou TH, Dang WM, Ma YT, Hu CQ, Wang N, Zhang GY, Wang G, Shi C, Zhang H, Guo B, et al. 2018. Clinical efficacy, onset time and safety of bright light therapy in acute bipolar depression as an adjunctive therapy: A randomized controlled trial. J Affect Disord. 227:90–96. doi:10.1016/j.jad.2017.09.038.

Addiction: Beyond dopamine reward circuitry

Nora D. Volkow[a,b,1], Gene-Jack Wang[c], Joanna S. Fowler[c], Dardo Tomasi[b], and Frank Telang[b]

[a]National Institute on Drug Abuse, National Institutes of Health, Bethesda, MD 20892; [b]National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD 20892; and [c]Medical Department, Brookhaven National Laboratory, Upton, NY 11973

Edited by Donald W. Pfaff, The Rockefeller University, New York, NY, and approved November 9, 2010 (received for review August 31, 2010)

Dopamine (DA) is considered crucial for the rewarding effects of drugs of abuse, but its role in addiction is much less clear. This review focuses on studies that used PET to characterize the brain DA system in addicted subjects. These studies have corroborated in humans the relevance of drug-induced fast DA increases in striatum [including nucleus accumbens (NAc)] in their rewarding effects but have unexpectedly shown that in addicted subjects, drug-induced DA increases (as well as their subjective reinforcing effects) are markedly blunted compared with controls. In contrast, addicted subjects show significant DA increases in striatum in response to drug-conditioned cues that are associated with self-reports of drug craving and appear to be of a greater magnitude than the DA responses to the drug. We postulate that the discrepancy between the expectation for the drug effects (conditioned responses) and the blunted pharmacological effects maintains drug taking in an attempt to achieve the expected reward. Also, whether tested during early or protracted withdrawal, addicted subjects show lower levels of D2 receptors in striatum (including NAc), which are associated with decreases in baseline activity in frontal brain regions implicated in salience attribution (orbitofrontal cortex) and inhibitory control (anterior cingulate gyrus), whose disruption results in compulsivity and impulsivity. These results point to an imbalance between dopaminergic circuits that underlie reward and conditioning and those that underlie executive function (emotional control and decision making), which we postulate contributes to the compulsive drug use and loss of control in addiction.

prefrontal cortex | dorsal striatum | substance use disorders | stimulant drugs | brain imaging

Drugs of abuse (including alcohol) are inherently rewarding, which is why they are consumed by humans or self-administered by laboratory animals (1). Only a small percentage of individuals exposed to drugs will become addicted, that is, shift from controlled drug use to compulsive drug use with loss of control over intake despite adverse consequences, however (2). Factors that determine who becomes addicted include genetic (50% of risk), developmental (risk is higher in adolescence), and environmental (e.g., drug access, stress) factors (2).

The mesolimbic dopamine (DA) pathway [DA cells in ventral tegmental area (VTA) projecting into nucleus accumbens (NAc)] seems to be crucial for drug reward (1). Other DA pathways [mesostriatal (DA cells in substantia nigra {SN} projecting into dorsal striatum) and mesocortical (DA cells in VTA projecting into frontal cortex)] are now also recognized to contribute to drug reward and addiction (1). The mode of DA cell firing also differently modulates the rewarding and conditioning effects, of drugs (predominantly phasic DA cell firing) vs. the changes in executive function that occur in addiction (predominantly tonic DA cell firing) (3, 4).

This review summarizes studies that used PET to evaluate DA's role in drug reward and addiction. These findings show that addiction affects not only the DA reward circuit but circuits involved with conditioning/habits, motivation, and executive functions (inhibitory control, salience attribution, and decision making). Other neurotransmitters (and neuropeptides) are involved with drug reward (i.e., cannabinoids, opioids) and with the neuroadaptations from repeated drug use (i.e., glutamate, opioids, GABA, corticotropic-releasing factor). These are not discussed here (except for glutamate), but several reviews address them (5, 6).

DA and Acute Drug Reward

All drugs that can lead to addiction increase DA in NAc, which is achieved through their interaction with different molecular targets by the various drug classes (6) (Table 1). In humans, PET studies have shown that several drugs [stimulants (7, 8), nicotine (9), alcohol (10), and marijuana (11)] increase DA in dorsal and ventral striatum (where NAc is located). These studies used a radiotracer that binds to DA D2 receptors (D2Rs) but only when these are not occupied by DA (i.e., [^{11}C]raclopride). By comparing binding after placebo and after the drug, these studies estimate the decreases in D2R availability induced by the drug, which are proportional to DA increases (12). Most studies have reported that participants who display the greatest DA increases with the drug also report the most intense "high" or "euphoria" (reviewed ref. 13).

PET studies have also shown that the speed with which a drug enters and leaves the brain (pharmacokinetic profile) is crucial for its reinforcing effects. Specifically, PET studies of brain pharmacokinetics of drugs labeled with positron emitters show that peak levels in human brain are reached within 10 min after i.v. administration and that this fast drug uptake is associated with the high (13) (Fig. 1). Indeed, for an equivalent level of cocaine reaching the brain (assessed as equivalent level of DA transporter blockade), when cocaine entered the brain rapidly (smoked and i.v. administration), it elicited a more intense high than when it entered the brain more slowly (snorted) (14). This is consistent with preclinical studies showing that the faster the drug's entry into the brain, the stronger are its reinforcing effects (15). This probably reflects the fact that abrupt and large DA increases triggered by drugs mimic the fast and large DA increases associated with phasic DA firing that are associated in the brain with conveying information about reward and saliency (16).

Drug-induced DA increases in NAc occur in nonaddicted as well as addicted subjects, which raises the question of how they relate to addiction.

To start with, there is increasing evidence that DA's role in reinforcement is more complex than just coding for reward per se (hedonic pleasure) and that stimuli that induce fast and large DA increases also trigger conditioned responses and elicit incentive motivation to procure them (17). Through conditioning, a neutral stimulus that is linked with the reinforcer (i.e., natural reinforcer, drug) acquires the ability by

This article arises from the Sackler Lecture, "Addiction: Conflict Between Brain Circuits," presented by Nora Volkow on June 11 at the AAAS Auditorium in Washington, DC. The lecture opened the Arthur M. Sackler Colloquium of the National Academy of Sciences, on "Quantification of Behavior." The complete program and audio files of most presentations are available on the NAS Web site at www.nasonline.org/quantification. See all papers from this colloquium in supplement 3 of volume 108.

Author contributions: N.D.V., G.-J.W., and J.S.F. designed research; N.D.V., G.-J.W., J.S.F., D.T., and F.T. performed research; N.D.V., G.-J.W., J.S.F., D.T., and F.T. analyzed data; and N.D.V. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: nvolkow@nida.nih.gov.

Downloaded by guest on February 17, 2022

Table 1. Main classes of drugs of abuse, their main molecular targets, and some of the mechanism(s) by which they increase DA in NAc (44)

Drug	Target	Mechanism for DA increases
Stimulant drugs (cocaine, amphetamine, methamphetamine)	DAT	Blocks DAT on the terminals of DA projecting neurons from VTA to NAc (cocaine) or releases DA from the vesicles of DA terminals (methamphetamine, amphetamine)
Opioids (heroin, opioid analgesics)	MOR	Disinhibits VTA DA neurons by inhibiting GABA interneurons that contain MOR in the VTA or directly activates NAc neurons that contain MOR
Nicotine (cigarettes and other tobacco products)	Nicotinic receptors (predominantly $\alpha4\beta2$ subtype)	Directly activates VTA DA neurons by stimulating their nicotine receptors and indirectly activates them by stimulating the nicotine receptors in glutamatergic terminals to VTA DA neurons
Alcohol and inhalants	Multiple targets, including GABA and glutamate receptors	Facilitates GABAergic neurotransmission, which may disinhibit VTA DA neurons from GABA interneurons or may inhibit glutamate terminals that regulate DA release in Nac
Cannabinoids (marihuana)	Cannabinoid CB1 receptors	Regulates dopaminergic signaling through CB1R in NAc neurons and in GABA and glutamate terminals to NAc

DAT, DA transporter; MOR, μ-opioid receptor.

itself to increase DA in striatum (including NAc) in anticipation of the reward, and this is associated with drug seeking (17). In animals trained to expect a natural reinforcer (food) when exposed to a conditioned stimulus (CS), the DA neurons stop responding to the primary reinforcer and, instead, respond to the CS (16). The extent to which a similar process occurs in response to drugs of abuse is unclear, however, because drugs, through their pharmacological actions, can directly activate DA neurons

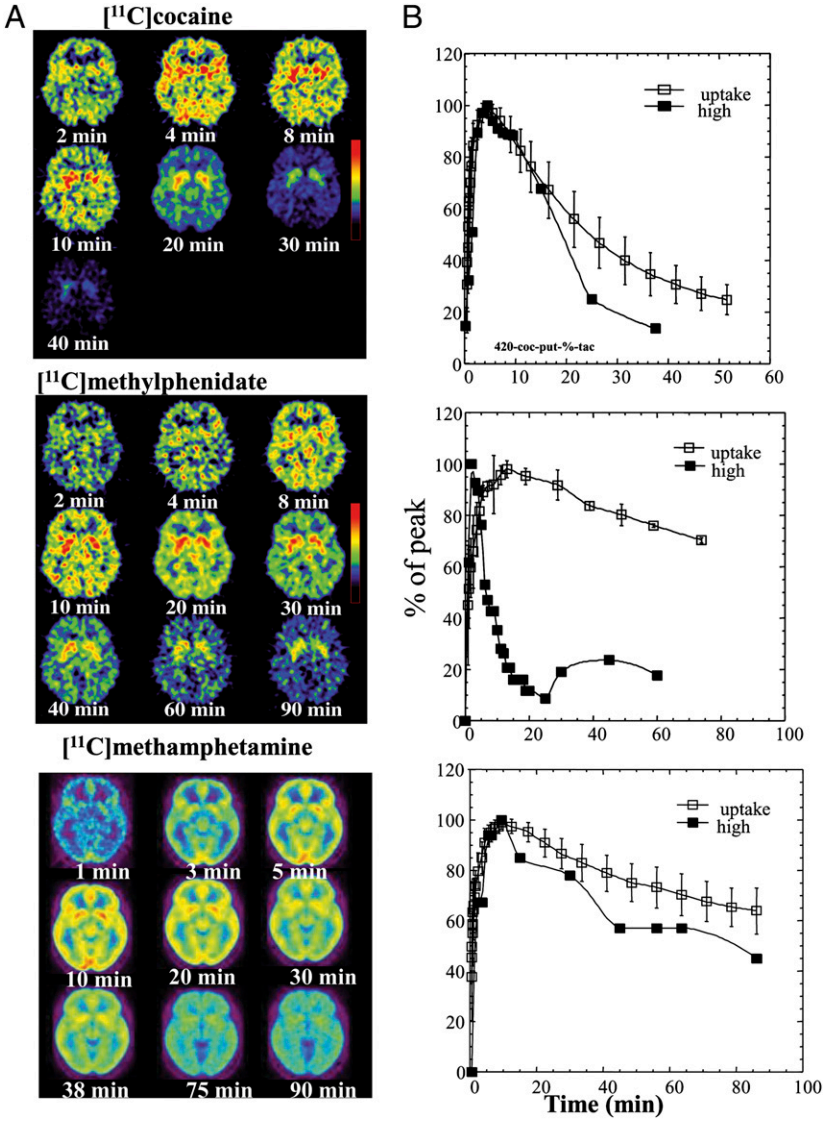


Fig. 1. Pharmacokinetics of stimulant drugs in the human brain and relationship to the "high." (*A*) Axial brain images of the distribution of [^{11}C]cocaine, [^{11}C]MP, and [^{11}C]methamphetamine at different times (minutes) after its administration. (*B*) Time activity curve for the concentration of [^{11}C]cocaine, [^{11}C]MP, and [^{11}C]methamphetamine in striatum alongside the temporal course for the high experienced after i.v. administration of these drugs. Note that the fast brain uptake for these drugs corresponds to the temporal course of the high, which suggests that the high is associated with the "rate of DA increases." In contrast, their clearance shows a correspondence with the high for cocaine and for methamphetamine but not for MP. The difference between MP and cocaine may reflect the differences in their rate of clearance and that between MP and methamphetamine may reflect their different mechanisms of action. Specifically, because MP and cocaine increase DA by blocking DA transporters, the DA increases are terminated by autoreceptor activation, which inhibits DA release. For cocaine, its fast rate of clearance (20-min half-life in brain) results in short-lasting autoreceptor activation, whereas for MP, its slower clearance (60-min half-life) results in long-lasting inhibition of DA release by autoreceptors, which terminates the high even though the drug is still in the brain. In contrast, methamphetamine, which is a DA releaser, is not sensitive to autoreceptor activation; thus, DA increases are not terminated by this mechanism, accounting for the longer lasting duration of the high. Modified from ref. 18. Copyright (1995) American Medical Association. All rights reserved. Reprinted from ref. 43. Copyright (2008), with permission from Elsevier.

Downloaded by guest on February 17, 2022

(i.e., nicotine) or increase DA release (i.e., amphetamine) (Table 1).

To answer this, we compared DA increases induced by the stimulant drug methylphenidate (MP) between cocaine-addicted subjects and controls. Like cocaine, MP increases DA by blocking DA transporters; both drugs have a similar distribution in human brain and have similar behavioral effects when given i.v. (18). In detoxified cocaine-addicted subjects (n = 20, detoxified 3–6 wk), we showed marked attenuation of MP-induced DA increases in striatum (50% lower) and of the increases in self-reports of high, compared with non–drug-abusing controls (n = 23). Similar findings were reported after administration of i.v. amphetamine (another stimulant drug) in recently detoxified cocaine abusers (detoxified 2 wk), who also showed decreased DA release in striatum and attenuated self-reports of euphoria (19). Because a confound in these studies was the possibility that drug withdrawal accounted for the attenuated DA responses, we repeated this study in active cocaine-addicted subjects (n = 19, nondetoxified) (20). In active cocaine abusers, MP-induced DA changes did not differ from placebo and the DA changes were 80% lower than in controls (n = 24); the self-reports of high were also attenuated (Fig. 2). Marked blunting of striatal DA increases secondary to MP or to amphetamine has also been documented in detoxified alcoholics (reviewed in ref. 13). If, as is currently believed, drug-induced DA increases in NAc underlie drug reward, why would cocaine-addicted subjects, who show a marked attenuation of drug-induced DA increases, compulsively take the drug?

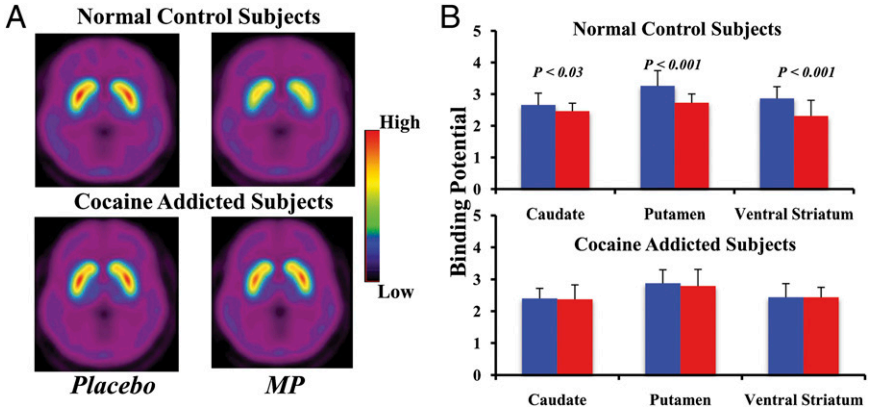


Fig. 2. DA changes induced by i.v. MP in controls and in active cocaine-addicted subjects. (*A*) Average nondisplaceable biding potential (BP_{ND}) images of [^{11}C]raclopride in active cocaine-addicted subjects (n = 19) and in controls (n = 24) tested after placebo and after i.v. MP. (*B*) D2R availability (BP_{ND}) in caudate, putamen, and ventral striatum after placebo (blue) and after MP (red) in controls and in cocaine-addicted subjects. MP reduced D2R in controls but not in cocaine-addicted subjects. Note that cocaine abusers show both decreases in baseline striatal D2R availability (placebo measure) and decreases in DA release when given i.v. MP (measured as decreases in D2R availability from baseline). Although one could question the extent to which the low striatal D2R availability in cocaine-addicted subject limits the ability to detect further decreases from MP, the fact that cocaine-addicted subjects show reductions in D2R availability when exposed to cocaine cues (Fig. 3) indicates that the attenuated effects of MP on [^{11}C] raclopride binding reflect decreased DA release.

DA and Conditioning to Drug Cues

The explanation may arise from the process of conditioning, which is one of the initial neuroadaptations that follow exposure to drugs and involves DA phasic signaling (predominantly D1Rs) and synaptic changes in NMDA and AMPA receptors (modulated by glutamate) (21, 22). These conditioned responses are believed to underlie the intense desire for the drug (craving) and the compulsive use that occurs when addicted subjects are exposed to drug cues.

To assess if drug conditioned cues would increase DA in cocaine-addicted subjects, we tested active cocaine-addicted subjects (n = 18) when subjects watched a neutral video (nature scenes) vs. when they watched a cocaine-cue video (scenes of subjects procuring and smoking cocaine) (23). Cocaine cues significantly increased DA in dorsal striatum, and the magnitude of this increase was correlated with the subjective experience of craving (Fig. 3); similar findings were reported by another laboratory (24). Subjects with the largest cue-induced DA increases in dorsal striatum also had the highest scores on measures of addiction severity. Because the dorsal striatum is implicated in habit learning, this association is likely to reflect the strengthening of habits as chronicity of addiction progresses. This suggests that a basic disruption in addiction might be DA-triggered conditioned responses that result in habits leading to compulsive drug consumption. Inasmuch as in cocaine-addicted subjects, the DA increases triggered by conditioned cues appear to be larger than those produced by a stimulant drug, this suggests that conditioned responses may drive the DA signaling that triggers and maintains the motivation to take the drug. To the extent that the drug (even when its DA-enhancing effects are attenuated) predicts reward, the act of its administration (e.g., injection, smoking) may become a conditioned cue and, as such, may increase DA. Thus, although drugs may initially lead to DA release in striatum (signaling reward), with repeated administration and as habits develop, there appears to be a shift in the DA increases from the drug to the CS, as reported for

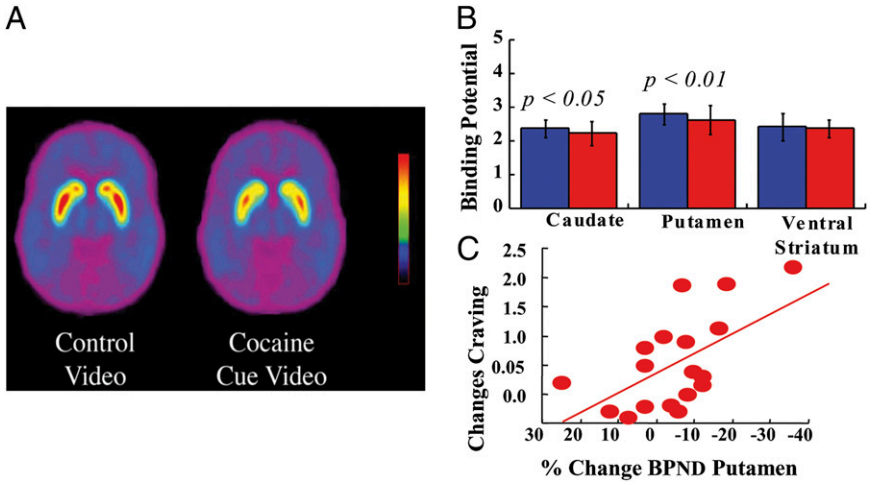


Fig. 3. DA changes induced by conditioned cues in active cocaine-addicted subjects. (*A*) Average nondisplaceable biding potential (BP_{ND}) images of [^{11}C]raclopride in cocaine-addicted subjects (n = 17) tested while viewing a neutral video (nature scenes) and while viewing a cocaine-cues video (subjects administering cocaine). (*B*) D2R availability (BP_{ND}) in caudate, putamen, and ventral striatum for the neutral video (blue) and the cocaine-cues video (red). The cocaine cues decreased D2R in caudate and putamen. (*C*) Correlations between changes in D2R (reflecting DA increases) and self-reports of cocaine craving induced by the cocaine-cues video. Modified from ref. 23.

Downloaded by guest on February 17, 2022

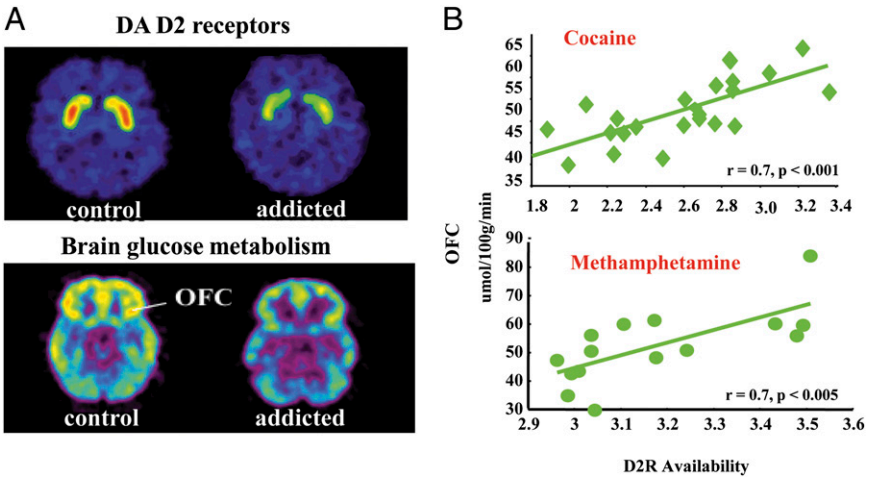


Fig. 4. Correlations between striatal D2R availability and metabolism in prefrontal brain regions. (*A*) Axial brain images for a control and for a cocaine-addicted subject for baseline images of D2R availability in striatum (obtained with [^{11}C]raclopride) and of brain glucose metabolism in OFC (obtained with [^{18}F]FDG). (*B*) Correlations between striatal D2R and metabolism in OFC in cocaine-addicted and methamphetamine-addicted subjects. Reprinted from ref. 13, Copyright (2009), with permission from Elsevier.

natural reinforcers (16). Preclinical studies have revealed that glutamatergic projections from prefrontal cortex into VTA/SN and NAc mediate these conditioned responses (5).

DA and Inhibitory Control in Addiction

The capacity to inhibit prepotent responses is likely to contribute to an individual's ability to restrain from taking drugs, and thus his or her vulnerability to addiction (25).

PET studies have shown that addicted subjects have significant reductions in D2R availability in striatum that persist months after protracted detoxification (reviewed in ref. 13). To investigate the functional significance of the striatal D2R reductions, we have assessed their relationship to baseline measures of brain glucose metabolism (marker of brain function). We have shown that reductions in striatal D2R are associated with decreased metabolism in orbitofrontal cortex (OFC), anterior cingulate gyrus (ACC), and dorsolateral prefrontal cortex (DLPFC) (26–28) (Fig. 4). Because OFC, CG, and DLPFC are involved with salience attribution, inhibitory control/emotion regulation, and decision making, we had postulated that their improper regulation by DA in addicted subjects could underlie the enhanced motivational value of drugs in their behavior and loss of control over drug intake (29). In addition, because impairments in OFC and ACC are associated with compulsive behaviors and impulsivity, we postulated that DA's impaired modulation of these regions could underlie the compulsive and impulsive drug intake seen in addiction (30, 31). Indeed, in methamphetamine abusers, low striatal D2R was

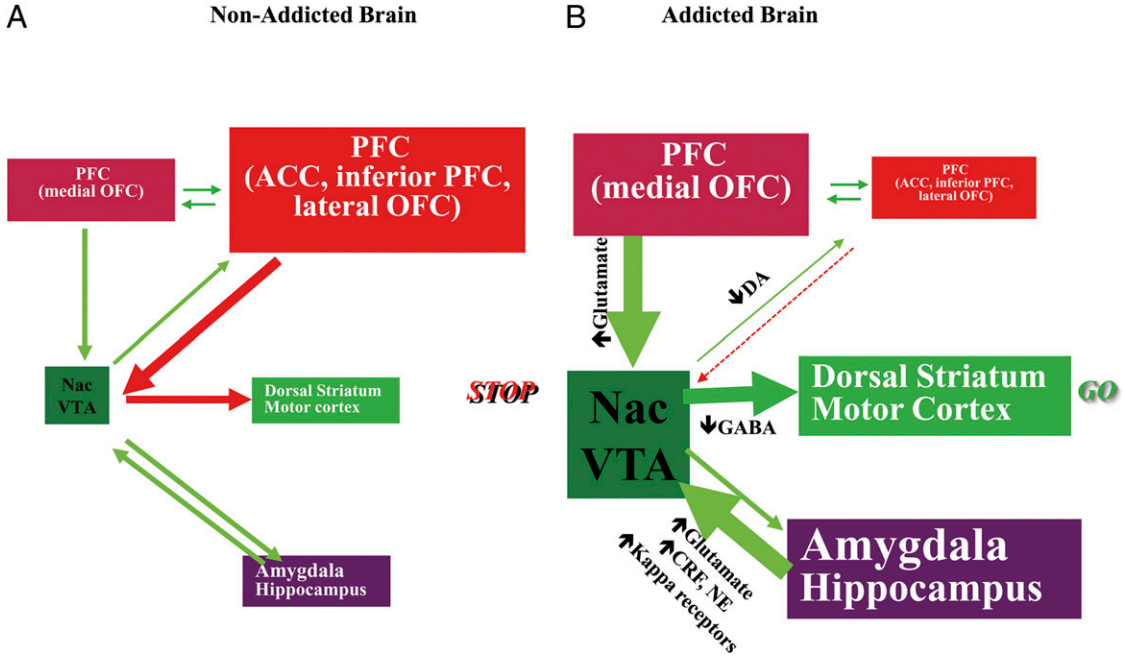


Fig. 5. Model proposing a network of interacting circuits underlying addiction: reward (nucleus accumbens, VTA, and ventral pallidum), conditioning/memory (amygdala, medial OFC for attribution of saliency, hippocampus, and dorsal striatum for habits), executive control (DLPFC, ACC, inferior frontal cortex, and lateral OFC), and motivation/drive (medial OFC for attribution of saliency, ventral ACC, VTA, SN, dorsal striatum, and motor cortex). Nac, nucleus accumbens. (*A*) When these circuits are balanced, this results in proper inhibitory control and decision making. (*B*) During addiction, when the enhanced expectation value of the drug in the reward, motivation, and memory circuits overcomes the control circuit, this favors a positive-feedback loop initiated by the consumption of the drug and perpetuated by the enhanced activation of the motivation/drive and memory circuits. These circuits also interact with circuits involved in mood regulation, including stress reactivity (which involves the amygdala and hypothalamus) and interoception (which involves the insula and ACC and contributes to awareness of craving). Several neurotransmitters are implicated in these neuroadaptations, including glutamate, GABA, norepinephrine, corticotropic-releasing factor, and opioid receptors. CRF, corticotropic-releasing factor; NE, norepinephrine. Modified with permission from ref. 35; permission conveyed through Copyright Clearance Center, Inc.

Downloaded by guest on February 17, 2022

associated with impulsivity (32), and low striatal D2R was associated with impulsivity and predicted compulsive cocaine administration in rodents (33).

It is also possible that the initial vulnerability for drug use occurs in prefrontal regions and that repeated drug use triggers the decreases in striatal D2R. Indeed, in a study done in subjects who, despite having a high risk for alcoholism (positive family history of alcoholism), were not alcoholics, we showed higher than normal striatal D2R availability that was associated with normal metabolism in OFC, ACC, and DLPFC (25). We interpreted this to suggest that normal prefrontal function may have protected these subjects from alcohol abuse.

DA and Motivation in Addiction

DA is also involved in motivation (i.e., vigor, persistence, effort toward the pursuit of reinforcing stimuli) through its regulation of several target regions, including NAc, ACC, OFC, DLPFC, amygdala, dorsal striatum, and ventral pallidum (34).

The enhanced motivation to procure drugs is a hallmark of addiction. Drug-addicted individuals will go to extreme behaviors to obtain drugs, even at the expense of seriously adverse consequences (2). Drug seeking and drug taking become their main motivational drives, which displace other activities (35). Thus, the addicted subject is aroused and motivated when seeking to procure the drug but tends to be withdrawn and apathetic when exposed to non–drug-related activities. This shift has been studied by comparing the brain activation patterns occurring with exposure to conditioned cues with those occurring without such cues.

In contrast to the decreases in prefrontal activity reported in detoxified cocaine abusers when not stimulated with drug or drug cues (reviewed in ref. 13), these prefrontal regions become activated when cocaine abusers are exposed to craving-inducing stimuli (either drugs or cues) (36–39). Similarly, cocaine abusers studied shortly after an episode of cocaine binging showed increased metabolic activity in OFC and ACC (also dorsal striatum) that was associated with craving (40).

Moreover, when we compared the response to i.v. MP between cocaine-addicted and nonaddicted subjects, we showed that MP increased metabolism in ventral ACC and medial OFC (an effect associated with craving) only in addicted subjects, whereas it decreased metabolism in these regions in nonaddicted subjects (41). This suggests that the activation of these prefrontal regions with drug exposure may be specific to addiction and associated with the enhanced desire for the drug. Moreover, in a subsequent study in which we prompted cocaine-addicted subjects to inhibit craving purposefully when exposed to drug cues, we showed that subjects who were successful in inhibiting craving decreased metabolism in medial OFC (processes motivational value of reinforcer) and NAc (predicts reward) (42).

These findings corroborate the involvement of OFC, ACC, and striatum in the enhanced motivation to procure the drug in addiction.

Systems Model of Addiction

As summarized above, several brain circuits are relevant in the neurobiology of addiction. Here, we highlighted four of these circuits: reward/saliency, motivation/drive, conditioning/habits, and inhibitory control/executive function (Fig. 5). The mood regulation circuit (contributes to regulation of stress reactivity) and the interoception circuit (contributes to awareness of drug craving and mood) also participate in addiction, but their involvement in the human brain has been much less investigated. Consequences of the disruption of these circuits are an enhanced motivational value of the drug (secondary to learned associations through conditioning and habits) at the expense of other reinforcers (secondary to decreased sensitivity of the reward circuit) and an impaired ability to inhibit the intentional actions associated with the strong desire to take the drug (secondary to impaired executive function) that result in compulsive drug taking in addiction (35).

Although it is likely that DA changes underlie some aberrant behaviors in addiction, it is also possible that some DA changes may reflect attempts to compensate for deficits in other neurotransmitters, particularly because DA is modulated by glutamate (GABA has been less investigated). Corticostriatal glutamatergic terminals are responsible for learning well-established behaviors and for changing these behaviors when they are no longer adaptive, and neuroadaptations in these projections (and in amygdalostriatal glutamate pathways) with repeated drug use (including impaired regulation of glutamate synaptic release) are implicated in the enhanced motivation for drug seeking that occurs in addiction (5). Impairments in glutamate-induced neuroplasticity with chronic drug exposure are also likely to be involved in the prefrontal function deficits reported in addicted individuals that result in impairments in inhibitory control and in inability to change maladaptive behaviors and to learn from the adverse consequences of drug use.

This model suggests a multipronged therapeutical approach to addiction to decrease the reinforcing properties of drugs, enhance the rewarding properties of natural reinforcers, inhibit conditioned-learned associations, enhance motivation for non–drug-related activities, and strengthen inhibitory control.

ACKNOWLEDGMENTS. We thank Linda Thomas for editorial assistance.

1. Wise RA (2009) Roles for nigrostriatal—not just mesocorticolimbic—dopamine in reward and addiction. *Trends Neurosci* 32:517–524.
2. Volkow N, Li TK (2005) The neuroscience of addiction. *Nat Neurosci* 8:1429–1430.
3. Wanat MJ, Willuhn I, Clark JJ, Phillips PE (2009) Phasic dopamine release in appetitive behaviors and drug addiction. *Curr Drug Abuse Rev* 2:195–213.
4. Grace AA (2000) The tonic/phasic model of dopamine system regulation and its implications for understanding alcohol and psychostimulant craving. *Addiction* 95(Suppl 2):S119–S128.
5. Kalivas PW (2009) The glutamate homeostasis hypothesis of addiction. *Nat Rev Neurosci* 10:561–572.
6. Koob GF (1992) Neural mechanisms of drug reinforcement. *Ann N Y Acad Sci* 654:171–191.
7. Volkow ND, et al. (1999) Reinforcing effects of psychostimulants in humans are associated with increases in brain dopamine and occupancy of D(2) receptors. *J Pharmacol Exp Ther* 291:409–415.
8. Drevets WC, et al. (2001) Amphetamine-induced dopamine release in human ventral striatum correlates with euphoria. *Biol Psychiatry* 49:81–96.
9. Brody AL, et al. (2009) Ventral striatal dopamine release in response to smoking a regular vs a denicotinized cigarette. *Neuropsychopharmacology* 34:282–289.
10. Boileau I, et al. (2003) Alcohol promotes dopamine release in the human nucleus accumbens. *Synapse* 49: 226–231.
11. Bossong MG, et al. (2009) Delta 9-tetrahydrocannabinol induces dopamine release in the human striatum. *Neuropsychopharmacology* 34:759–766.
12. Breier A, et al. (1997) Schizophrenia is associated with elevated amphetamine-induced synaptic dopamine concentrations: Evidence from a novel positron emission tomography method. *Proc Natl Acad Sci USA* 94: 2569–2574.
13. Volkow ND, Fowler JS, Wang GJ, Baler R, Telang F (2009) Imaging dopamine's role in drug abuse and addiction. *Neuropharmacology* 56(Suppl 1):3–8.
14. Volkow ND, et al. (2000) Effects of route of administration on cocaine induced dopamine transporter blockade in the human brain. *Life Sci* 67:1507–1515.
15. Balster RL, Schuster CR (1973) Fixed-interval schedule of cocaine reinforcement: Effect of dose and infusion duration. *J Exp Anal Behav* 20:119–129.
16. Schultz W (2010) Dopamine signals for reward value and risk: Basic and recent data. *Behav Brain Funct* 6: 24.
17. Owesson-White CA, et al. (2009) Neural encoding of cocaine-seeking behavior is coincident with phasic dopamine release in the accumbens core and shell. *Eur J Neurosci* 30:1117–1127.
18. Volkow ND, et al. (1995) Is methylphenidate like cocaine? Studies on their pharmacokinetics and distribution in the human brain. *Arch Gen Psychiatry* 52:456–463.
19. Martinez D, et al. (2007) Amphetamine-induced dopamine release: Markedly blunted in cocaine dependence and predictive of the choice to self-administer cocaine. *Am J Psychiatry* 164:622–629.
20. Wang G-J, et al. (2010) Decreased brain dopaminergic responses in active cocaine dependent subjects. *J Nucl Med* 51(Suppl 2):74.
21. Zweifel LS, et al. (2009) Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. *Proc Natl Acad Sci USA* 106:7281–7288.
22. Kauer JA, Malenka RC (2007) Synaptic plasticity and addiction. *Nat Rev Neurosci* 8:844–858.

Downloaded by guest on February 17, 2022

23. Volkow ND, et al. (2006) Cocaine cues and dopamine in dorsal striatum: Mechanism of craving in cocaine addiction. *J Neurosci* 26:6583–6588.
24. Wong DF, et al. (2006) Increased occupancy of dopamine receptors in human striatum during cue-elicited cocaine craving. *Neuropsychopharmacology* 31:2716–2727.
25. Volkow ND, et al. (2006) High levels of dopamine D2 receptors in unaffected members of alcoholic families: Possible protective factors. *Arch Gen Psychiatry* 63: 999–1008.
26. Volkow ND, et al. (1993) Decreased dopamine D2 receptor availability is associated with reduced frontal metabolism in cocaine abusers. *Synapse* 14:169–177.
27. Volkow ND, et al. (2001) Low level of brain dopamine D2 receptors in methamphetamine abusers: Association with metabolism in the orbitofrontal cortex. *Am J Psychiatry* 158:2015–2021.
28. Volkow ND, et al. (2007) Profound decreases in dopamine release in striatum in detoxified alcoholics: Possible orbitofrontal involvement. *J Neurosci* 27:12700–12706.
29. Volkow ND, Fowler JS (2000) Addiction, a disease of compulsion and drive: Involvement of the orbitofrontal cortex. *Cereb Cortex* 10:318–325.
30. Volkow ND, Ding YS, Fowler JS, Wang GJ (1996) Cocaine addiction: Hypothesis derived from imaging studies with PET. *J Addict Dis* 15:55–71.
31. Goldstein RZ, Volkow ND (2002) Drug addiction and its underlying neurobiological basis: Neuroimaging evidence for the involvement of the frontal cortex. *Am J Psychiatry* 159:1642–1652.
32. Lee B, et al. (2009) Striatal dopamine d2/d3 receptor availability is reduced in methamphetamine dependence and is linked to impulsivity. *J Neurosci* 29:14734–14740.
33. Everitt BJ, et al. (2008) Review. Neural mechanisms underlying the vulnerability to develop compulsive drug-seeking habits and addiction. *Philos Trans R Soc Lond B Biol Sci* 363:3125–3135.
34. Salamone JD, Correa M, Farrar A, Mingote SM (2007) Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology (Berl)* 191:461–482.
35. Volkow ND, Fowler JS, Wang GJ (2003) The addicted human brain: Insights from imaging studies. *J Clin Invest* 111:1444–1451.
36. Volkow ND, et al. (1999) Association of methylphenidate-induced craving with changes in right striato-orbitofrontal metabolism in cocaine abusers: Implications in addiction. *Am J Psychiatry* 156:19–26.
37. Wang G-J, et al. (1999) Regional brain metabolic activation during craving elicited by recall of previous drug experiences. *Life Sci* 64:775–784.
38. Grant S, et al. (1996) Activation of memory circuits during cue-elicited cocaine craving. *Proc Natl Acad Sci USA* 93:12040–12045.
39. Daglish MR, Nutt DJ (2003) Brain imaging studies in human addicts. *Eur Neuropsychopharmacol* 13:453–458.
40. Volkow ND, et al. (1991) Changes in brain glucose metabolism in cocaine dependence and withdrawal. *Am J Psychiatry* 148:621–626.
41. Volkow ND, et al. (2005) Activation of orbital and medial prefrontal cortex by methylphenidate in cocaine-addicted subjects but not in controls: Relevance to addiction. *J Neurosci* 25:3932–3939.
42. Volkow ND, et al. (2010) Cognitive control of drug craving inhibits brain reward regions in cocaine abusers. *NeuroImage* 49:2536–2543.
43. Fowler JS, et al. (2008) Fast uptake and long-lasting binding of methamphetamine in the human brain: Comparison with cocaine. *NeuroImage* 43:756–763.
44. Nestler EJ, et al. (2005) Is there a common molecular pathway for addiction? *Nat Neurosci* 8:1445–1449.

Downloaded by guest on February 17, 2022

Economics and Human Biology 8 (2010) 233–241

Contents lists available at ScienceDirect

Economics and Human Biology

journal homepage: http://www.elsevier.com/locate/ehb

Brainstorm: Occupational choice, bipolar illness and creativity

Carol Horton Tremblay *, Shawna Grosskopf, Ke Yang [1]

Department of Economics, Oregon State University, Corvallis, OR 97331-3612, USA

ARTICLE INFO

Keywords:
Bipolar illness
Mental health
Occupation
Creativity

ABSTRACT

Although economists have analyzed earnings, unemployment, and labor force participation for those with bipolar illness, occupational choice has yet to be explored. Psychological and medical studies often suggest an association between bipolar illness and creative achievement, but they tend to focus on eminent figures, case studies, or small samples. We seek to examine occupational creativity of non-eminent individuals with bipolar disorder. We use Epidemiologic Catchment Area data to estimate a multinomial logit model matched to an index of occupational creativity. Those with bipolar illness appear to be disproportionately concentrated in the most creative occupational category. Nonparametric kernel density estimates reveal that the densities of the occupational creativity variable for the bipolar and non-bipolar individuals significantly differ in the ECA data, and suggest that the probability of engaging in creative activities on the job is higher for bipolar than non-bipolar workers.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The association between creativity and bipolar illness is an age-old controversy.[2] Bipolar illness is characterized by pronounced mood swings—alternating periods of elation, normalcy, and despondency, which can be accompanied by hallucinations or delusions during the peaks and troughs of the cycles. Although we might expect bipolar illness to hamper occupational and artistic achievement, the creativity and energy characteristic of persons with mania may actually improve some aspects of their labor market outcomes (Frank and McGuire, 2000).

Estimates of prevalence of bipolar illness fall around 1.2% of the population (Weissman et al., 1991).[3] With respect to the labor market, the employment of the bipolar population is approximately 71% compared to about 82% for the general population, and those with bipolar illness earn about 57% of the earnings of others, all else equal (Ettner et al., 1997). In contrast to poorer employment and earnings outcomes in the Ettner et al. (1997) study, Marcotte et al. (2000) find that bipolar status in the past year does not significantly affect income or employment for men or women, but that length of time with a bipolar

* Corresponding author at: Department of Economics, Oregon State University, Ballard Extension Hall 303, Corvallis, OR 97331-3612, USA. Tel.: +1 541 737 1468; fax: +1 541 737 5917.
E-mail address: ctremblay@oregonstate.edu (C.H. Tremblay).
[1] Present address: Barney School of Business at University of Hartford, 200 Bloomfield Avenue, West Hartford, CT 06117, USA.

[2] See Jamison (1993) for a thorough history and literature review of the topic.

[3] This figure is controversial, however. Some psychiatrists claim that bipolar illness is not a discrete phenomenon, but is better characterized as a bipolar spectrum. The fraction of the general population with bipolar spectrum disorder has been estimated at 4.5% in the National Comorbidity Survey Replication (Merikangas et al., 2007) and at 6.4% in the Epidemiological Catchment Area survey (Judd and Akiskal, 2003). At the same time, Zimmerman (2008) find that bipolar disorder is often overdiagnosed by health care professionals. In our work the diagnoses are made based on the National Institute of Mental Health Diagnostic Interview Schedule. There appears to be no difference in the prevalence of bipolar I disorder by sex, race, or ethnicity, and the typical age of onset is 20 years (APA, 2000).

1570-677X/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ehb.2010.01.001

disorder positively and significantly affects employment for women. Further, the lifetime number of bipolar episodes appears to significantly increase income for men. These mixed findings for the bipolar population suggest a need for further study.

None of the economic studies to date have addressed the issue of occupational outcomes for those with bipolar illnesses. Yet there is a substantial body of research on the possible link between creative and artistic occupations and bipolar illness in the medical and behavioral sciences literature. About 75% of these studies find evidence of such a link (Jamison, 1993). Much of this research derives from case studies, biographies of prominent historical figures in the arts, or diagnostic and psychological studies of living writers, artists, and composers. Although rich and informative in their own right, these studies suffer from small sample sizes or from selection bias. In contrast to the case study approach, we employ a population based data set to contribute to the literature on bipolar disorder and creativity as well as the labor market literature on occupational outcomes.

The clinical definition of mania may help in understanding the link between bipolar illness and creativity. Depression characterizes both unipolar and bipolar illness; thus it is an episode of mania that distinguishes bipolar illness (or bipolar I disorder) from unipolar illness. Generally speaking, a manic episode entails "a distinct period of abnormally and persistently elevated, expansive or irritable mood," which may alternate with depressive mood. A manic episode is characterized by at least three of the following symptoms: increased activity, talkativeness, flight of ideas, inflated self-esteem, decreased need for sleep, distractibility, and excessive involvement in risky activities. Some of these symptoms encourage creativity, for example "flight of ideas" and "talkativeness" (which might facilitate rhyming and verse). Increased activity and decreased need for sleep may bolster productivity in a number of occupations.

Another explanation is that there may be a genetic tie between bipolar disorder and creativity. Richards et al. (1988) find that unafflicted close relatives of bipolar patients tend to be more creative than the general population. In addition, Kèri (2009) studied the genomic DNA of 200 healthy, high intellectual achievement individuals, and found that those who carried a genotype related to psychosis risk had the highest creativity levels.[4] Some evolutionary psychologists argue that the positive benefits of creativity might account for the persistence of bipolar disorder in the gene pool in spite of the negative aspects of the illness, although there appears to be some controversy over the reasons for that persistence (Keller and Miller, 2006; Miller and Tal, 2007). It may be that the genetic complexities of mental illness have prevented its elimination by natural selection. For example, individual genes contributing to mental illness may not be harmful alone and may tie to creativity, but when in combination with other genes result in mental illness.

We do not presume to disentangle potential causal factors, but attempt to establish whether or not there is an association between bipolar illness and occupational creativity. Unlike prior studies, we bring a different kind of data, traditional methods from labor economics, as well as nonparametric statistical techniques to bear on this question. Rather than taking a case study approach, we focus on the non-eminent population and use a large population-based sample with automated implementation of the diagnostic criteria to distinguish bipolars from the rest of the sample (the Epidemiologic Catchment Area (ECA) data set).[5] To see if those with and those without bipolar illness have the same occupational outcomes, we use a multinomial logit model of occupational outcomes.

Once differences in occupational outcomes are evaluated, we investigate whether bipolars are concentrated in more creative occupations based on a measure from the *Dictionary of Occupational Titles*. To test for differences in this measure for bipolars we turn to nonparametric methods, since the distribution of the occupational creativity variable is highly skewed, rendering parametric regression analysis inappropriate. Instead we estimate separate nonparametric densities for bipolar and non-bipolar individuals, and test for differences in the distributions.

2. Data and descriptive analysis

The main dataset we utilize is the Epidemiologic Catchment Area Study (ECA), collected by researchers at Yale, Johns Hopkins, Washington, Duke and University of California at Los Angeles in collaboration with the National Institute of Mental Health (NIMH; U.S. Department of Health and Human Services, 1991). Data were obtained by personal interviews of 20,861 adults residing in the towns of these universities: New Haven, Baltimore, St. Louis, Durham and Los Angeles. Respondents were selected using multistage probability sampling. We use the second of two waves of the survey, which was collected in 1981–1985.[6]

To determine if an individual meets the criteria for having a particular mental illness, the ECA interviewers elicited responses to the NIMH Diagnostic Interview Schedule (DIS). Questions are designed to identify symptoms corresponding to criteria in the Diagnostic and Statistical Manual of Mental Disorders, third edition (DSM-III; American Psychiatric Association, 1980). Once the DIS responses are entered into a data file, the DSM-III criteria are operationalized by computer, and diagnoses are

[4] Psychosis refers to the experience of hallucinations and/or delusions (Goodwin and Jamison, 2007: 431; APA, 2000: 414).

[5] We also considered using data from the Mental Health Supplement (MHS) of the National Health Interview Survey. This is also a population-based sample, but with self reported diagnosis of bipolar illness in the last 12 months rather than lifetime as in the ECA. This is a less inclusive definition than that in the ECA sample, as reflected in the small percentage, 0.16, out of 53,441 respondents and more likely to include those with more acute symptoms than in the ECA. We focus instead on the ECA data.

[6] Although the sample is dated, the relationship between bipolar illness and creativity is not likely to change over time. The measure of occupational creativity, discussed below, corresponds to the 1980 occupational codes.

Table 1
Occupational distributions by manic episode status (%).

Occupation	No manic episode	Bipolar (DSM-III)	Bipolar (DSM-III-R)
Professional	20.59	27.38	30.43
Technical	29.54	26.19	21.74
Services	17.27	23.81	26.09
Crafts	10.86	7.14	4.35
Operators	21.73	15.48	17.39
Total percent	100.00	100.00	100.00
Total count	13,570	84	46

Note: Professional includes managerial occupations; technical includes sales and clerical; operators includes fabricators and laborers. DSM-III: Prior episode defined over lifetime. DSM-III-R: Same, but must also have been hospitalized or markedly occupationally impaired. *Source*: Epidemiologic Catchment Area survey, U.S. Department of Health and Human Services (1991), URL: http://dx.doi.org/10.3886/ICPSR08993.

Table 2
Occupational creativity index by occupation.

Occupation	Mean creativity	Standard error	Median creativity	*N*
Professional	10.14	0.40	2.00	2814
Technical	0.95	0.03	0.17	4029
Services	3.47	0.33	0.05	2374
Crafts	0.66	0.03	0.05	1482
Operators	0.45	0.04	0.02	2964
F	176.92			
Prob-F	<0.0001			
χ^2			2800	
Prob-χ^2			<0.0001	

Note: Index is from 0 to 100 with 0 the least creative. Professional includes managerial occupations; technical includes sales and clerical; operators includes fabricators and laborers.

generated for mania as well as a number of other mental illnesses. There are 84 individuals (that is, 0.62%) in the sample with information on occupation who have a DSM-III manic episode diagnosis.

We also consider an alternative diagnostic criteria, the 1987 revised version of the DSM-III, abbreviated DSM-III-R. In addition to the mania diagnosis requirements in the DSM-III, the DSM-III-R includes the criterion of hospitalization or marked impairment in occupational functioning or usual social activities. Excluding those who do not meet this severity criterion, reduces the number of individuals with a DSM-III-R manic episode diagnosis to a total of 46 or 0.34% of the sample.[7]

The ECA sample also contains occupational codes. The occupation variable applies to those currently and previously employed.[8] Table 1 lists the percent of observations in each major occupational group by manic episode status for the two forms of bipolar diagnosis in the ECA sample (DSM-III and DSM-III-R). There is a greater representation of those with a bipolar I diagnosis than the rest of the population in the professional and managerial occupations and services based on the sample data. Bipolars are relatively 'overrepresented' in the services. Likelihood ratio chi-squared tests for differences in occupation distributions by manic status show no significant difference. The probability values for the chi-square statistics are 0.15 for DSM-III criteria and 0.11 for DSM-III-R.

The occupational segregation index (Duncan and Duncan Index of Occupational Dissimilarity) corresponding to the numbers in Table 1 is 13.3, indicating that identical occupation distributions for bipolars and those with no history of mania would arise if 13.3% of either group changed occupations for bipolar DSM-IIIs.[9] To put this figure in perspective, note that it equals the value of the occupational segregation index for African and white American women in 1995 (Blau et al., 1998). The estimated occupational segregation index for DSM-III-R bipolars relative to non-bipolars is 18.7%, higher than for the bipolars diagnosed under the less restrictive DSM-III criteria.

To measure creativity for each individual, we use occupation creativity scores from England and Kilbourne (1988) matched to individuals by 3-digit occupation codes.[10] The occupational creativity measure represents the percentage of employees in a particular occupation

[7] There is a more recent diagnostic and statistical manual of mental disorders, the DSM-IV-TR. In this as well as the DSM-III and DSM-III-R, there is a list of factors that can cause "manic-like episodes" such as substance abuse or a medical condition like multiple sclerosis. When these factors precede symptoms of mania, a diagnosis of mania is not made. A primary change in criteria in the DSM-IV-TR, is that it expands and clarifies the list of factors that preclude a manic episode diagnosis to include factors such as medication and antidepressant light therapy treatment. Light therapy is such a recent tool, we do not believe that it is an issue in our data set. The DSM-IV-TR also lists antidepressant medication and electroconvulsive therapy, and although they are not listed in the DSM-III, they are in the DSM-III-R. For other differences in the manic episode criteria for the three manuals, see American Psychiatric Association (1980, 1987, 2000). The list of symptoms of a manic episode is virtually the same for the DSM-III, DSM-III-R, and DSM-IV-TR. To follow up on the medication and substance use issues we examine a model accounting for alcohol and drugs below.

[8] The occupation recorded for those not currently employed is most recent occupation.

[9] The occupational segregation index for bipolar and non-bipolar workers is given by: $D = (1/2)\sum_i |(B_i/B) - (NB_i/NB)|$ where i indexes occupation, B_i is the number of bipolar workers employed in occupation i, B is the number of bipolar workers in all occupations, NB_i is the number of non-bipolar workers employed in occupation i, and NB is the number of non-bipolar workers in all occupations. The value of the index ranges from 0 to 1, with 0 indicating no occupational segregation and 1 indicating complete segregation (Blau et al., 1998).

[10] The occupational creativity variable is derived from a measure of creativity from the Fourth Edition of the Dictionary of Occupational Titles (DOT) produced by the U.S. Department of Labor. The DOT contains extensive job requirement and characteristic information at a highly detailed level of occupation. The creativity variable reported by the DOT takes one of three values for the requirements of each detailed occupation: (−1)—preference for activities of a routine, concrete, organized nature; (0)—neither pole relevant to occupation; (1)—preference for activities of an abstract and creative nature. It is the latter value in which we are interested. The DOT data were merged with Census Bureau data by 3-digit occupational code, and the percentage of workers in the aggregated occupational group requiring a preference for abstract and creative activities was constructed. This is the occupational creativity variable we use here.

who engage in abstract and creative activities and ranges from 0 to 100.[11]

While Table 1 highlights the relationship between occupations and bipolar status, Table 2 focuses on the creativity levels of occupations, regardless of bipolar status. Table 2 shows the means, standard errors, and median values of the occupational creativity index by occupation group for all workers. Professional and managerial occupations have the highest mean and median creativity levels relative to the other occupational groups, and as shown in Table 1, this occupational category employs an 'over-representation' of those with bipolar illness. Further, the professional and managerial occupational group includes all of the 3-digit occupations classified by Filer (1986) as "artistic."[12] In the services, where bipolars are also over-represented (Table 1), the extent of creativity is less clear. Differences in mean creativity levels are highly significant across occupational groups by an F-test (p-value < 0.0001) as well as pairwise t-tests ($p = 0.001$ or better for all pairs). Differences in medians across occupations are also highly significant (p-value < 0.0001 for the Pearson χ^2 test). Thus creativity varies significantly by occupation for all workers, but the distribution of occupations differs, though not significantly, for bipolars and non-bipolars.

Taking a more direct approach, we segment the sample by diagnosis status and compute the occupational creativity index. Table 3 shows that both the mean and median values of the occupational creativity index are higher for bipolar DSM-IIIs and DSM-III-Rs (means = 4.35 and 4.54, medians = 0.46 and 0.38) than for those not so afflicted (mean = 3.07, median = 0.12).[13] The mean and median values of the bipolar and non-bipolar groups are not significantly different from one another, with the exception of the marginally significant difference in medians for DSM-III bipolars and non-bipolars.[14]

Table 3
Occupational creativity index by diagnosis.

	Mean creativity	Standard error	Median creativity	*N*
Diagnosis				
Bipolar (DSM-III)	4.35	1.62 **	0.46 *	84
Bipolar (DSM-III-R)	4.54	2.20 **	0.38	46
Non-bipolar	3.07	0.10	0.12	13,888

* Significantly different from the non-bipolar group at 10% by Fisher's exact two-sided test and Pearson χ^2 test.

** Variance of creativity variable significantly different from non-bipolar group at 5% or better.

We explore one other measure of creativity from the psychology and career counseling literature: the Holland Artistic (A) Occupational Code (Gottfredson and Holland, 1996).[15] We merged the Holland codes to our sample by 3-digit occupational code, and we used the DSM-III criteria to distinguish bipolars from others in the merged sample. We find that the percent of individuals who are in occupations with an Artistic component is greater for the bipolar group, 10.71 percent, than for the non-bipolar group, 8.40 percent. The difference is not statistically significant which is likely due to the small sample of bipolar individuals, 75 in the non-artistic occupations and only 9 in the artistic occupations.[16]

Interestingly, the variances of the creativity index differ significantly for the non-bipolar group and each of the bipolar groups (p-value = 0.010 for DSM-III and 0.044 for DSM-III-R by a variance ratio F-test). It might be possible that the variability in the expression of the illness across individuals—for example, the length and frequency of episodes, the severity of mania, depression and psychosis, the extent of impairment—results in a greater variance in occupational creativity than among the rest of the population.

In the next section we turn to traditional empirical models to analyze occupational choice. Specifically we use multinomial logit estimation of occupational choice to

[11] Examples of abstract and creative activities include: painting, hairstyling, writing, music teaching, interpreting public opinion surveys in light of contemporary society, creating dramatic stage lighting, planning advertising campaigns, and diagnosing illness (U.S. Department of Labor, 1972).

[12] These include actors and directors; authors; dancers; designers working in the theater, motion pictures, or art museums; musicians and composers; painters, sculptors, craft artists, and artistic printmakers; photographers; postsecondary teachers of art, drama and music; and artists, performers and related workers not elsewhere classified.

[13] The large differences between mean and median values of the creativity index reflect the highly skewed distribution of this variable, which we address below with nonparametric methods.

[14] Fisher's exact two-sided p-value = 0.080 and the p-value for the Pearson $\chi^2 = 0.076$ for the difference in medians for DSM-III bipolars and non-bipolars. p-Values for differences in means for those who are not bipolar relative to either bipolar group are p-value = 0.22 for DSM-III and 0.25 for DSM-III-R diagnoses for one-tailed t-tests with unequal variances. The difference in medians for the 46 persons diagnosed by DSM-III-R criteria and others is not significant by either the Fisher or the Pearson tests ($p = 0.363$ and $p = 0.379$, respectively). We also conduct a Wilcoxon rank-sum (Mann–Whitney) test and obtain insignificant test statistics (p-value = 0.217 for the DSM-III and p-value 0.5844 for the DSM-III-R group). Nonparametric estimation is further discussed below.

[15] According to Gottfredson and Holland (p. 2), the Artistic occupational environment "requires innovation or creative ability. It rewards the display of imagination in artistic, literary, or musical accomplishments, and allows the expression of unconventional ideas or manners. Occupations classified as Artistic generally involve creative work in the arts: music, writing, performance, sculpture, or other relatively unstructured and intellectual endeavors." The Holland model classifies five additional occupational environments: Realistic (R, e.g. carpenter or truck operator), Investigative (I, e.g. psychologist or microbiologist), Social (S, e.g. counselor or clergy member), Enterprising (E, e.g. lawyer or retail store manager), and Conventional (C, e.g. production editor, bookkeeper). For seven different occupational coding systems, Gottfredson and Holland assign three of the six categories, in descending order of importance, to each occupation. For example, the code for economists is ISA, investigative rated most important, followed by social (which includes teaching, mentoring and concern for the welfare of others), and then artistic (likely for the creativity and innovation features). The statistics referred to above apply to all occupations that have an Artistic component (i.e., Axx, xAx, or xxA where x = R, I, S, E, C).

[16] The p-values for a one-tailed t-test, a Pearson χ^2 test, and a Fisher exact test are 0.2226, 0.5537 and 0.2733, respectively. We include the p-value for the t-test because it is so commonly used. The t-test is not technically appropriate in this case, however, because the variable, percent of individuals who are in occupations with an Artistic component, derives from a set of discrete variables and is not normally distributed.

Table 4
Multinomial logit marginal effect estimates of bipolar status on occupational outcomes.

	Professional	Technical	Services	Crafts	Operators	*N*	LR
Base model, full sample							
Bipolar (DSM-III)	0.05 (0.96)	−0.05 (0.97)	**0.09** (1.75)	−0.03 (1.12)	−0.07 (1.49)	13,622	3235
Bipolar (DSM-III-R)	0.08 (1.16)	**−0.11** (1.76)	0.11 (1.49)	**−0.05** (1.69)	−0.04 (0.58)	13,584	3216
Full sample, controlling for alcohol and drugs							
Bipolar (DSM-III)	0.07 (1.43)	−0.04 (0.77)	0.07 (1.60)	−0.03 (1.24)	**−0.07** (1.84)	13,622	3378
Bipolar (DSM-III-R)	**0.11** (1.68)	−0.09 (1.51)	0.08 (1.32)	**−0.06** (1.77)	−0.05 (0.88)	13,584	3265

Note: Base model does not control for alcohol/drugs, marital status or education. Professional includes managerial occupations; technical includes sales and clerical; operators includes fabricators and laborers. Absolute values of *t*-ratios are in parentheses. N = number of observations. LR = likelihood ratio; all LR values are significant at 1%. Coefficients in bold are significant at 10%.

determine if bipolar status significantly affects occupational choice. We then return to analyzing the relationship between bipolar illness and the occupational creativity index. We use nonparametric techniques to estimate kernel densities of the occupational creativity index for bipolars and non-bipolars and test for differences in the distributions.

3. Results of multivariate analysis

3.1. Multinomial logit occupation regressions

We estimate a multinomial logit model of the probability of employment in each occupational group as a function of manic episode status. Control variables in the base model include RACE = 1 if individual i is African-American (= 0 otherwise), FEM = 1 if i is female (= 0 otherwise), and AGE. Sets of dummy variables for year of interview and interview site are also included. Although it might be desirable to include education level in the regression, it was not collected at Johns Hopkins or Washington University. For the set of remaining observations from the Yale, Duke, and UCLA communities, the mean education level is significantly higher for bipolar I's than for the rest of the sample at 1 percent or better (12.77 compared to 11.52 for the DSM-III group and 12.97 versus 11.52 for the DSM-III-R group). Since the sample is limited, results controlling for education are discussed below with alternate specifications.

Multinomial logit estimates of marginal effects of the bipolar dummy variable for alternative occupation groups are presented in Table 4. These include results for the full sample with the base model (no education, no marital status, no alcohol or drug variables included). The results show that the marginal effect of the bipolar dummy variable for a manic episode is positive across the board in the most 'creative' occupations: professional and managerial and services, although the marginal effects are significant at the 10 percent or better level in only one case: DSM-III for services. The marginal effect is consistently negative for the other 'low creativity' occupations, and is significant at 10 percent in two cases in the base model: for technical, sales and clerical and for crafts for DSM-III-R in both cases.

Recall that in contrast to the DSM-III diagnosis of bipolar illness, DSM-III-R requires hospitalization or marked impairment in occupational functioning or usual social activities, which reduces the number of individuals classified as bipolar in the sample to 46. Note that if we wish to investigate the relationship between bipolar illness and occupational creativity, and if occupational impairment is used as a criterion for bipolar illness, we might expect to understate occupational creativity for the DSM-III-R classification. However, when the DSM-III variable is replaced by the DSM-III-R dummy variable for a mania diagnosis in the logit model, we see an increase in the absolute value of the bipolar marginal effect[17] suggesting even stronger selection of bipolars into relatively creative occupations.

The occupational impairment rule in the DSM-III-R diagnostic criteria has identified bipolars as more frequently employed in creative occupations with a higher mean (but lower median) level of occupational creativity, a greater degree of occupational segregation relative to the well, and somewhat stronger multinomial logit evidence linking the presence of the disease to creative occupations than the bipolars identified under the DSM-III rules. Thus, even with the potential bias this new screen may impose, there appears to be some evidence of a link between occupational outcomes, creativity and bipolar disorder.

3.2. Alternative specifications

We estimate a number of alternative specifications, and the substantive results are unchanged. We add a marital status dummy variable and then an education variable to the set of demographic variables in the base model. Although we do not report them here, (results available on request), the signs of the estimated marginal effects on the bipolar dummy variables are robust to these specification changes. For both full sample and 3-site sample, and both DSM-III and DSM-III-R criteria, the signs are the same as in the base model: positive for the professional and managerial occupations, negative for the technical, sales and clerical jobs; positive for the service jobs; negative for the crafts; and negative for the operators, fabricators and laborers.

We next consider another set of specifications which control for co-morbid psychiatric illnesses. Each model includes a dummy variable for schizophrenia, obsessive-compulsive disorder, phobia, panic disorder, anti-social personality or alcohol and drug abuse.[18] The signs on the

[17] With the exception of the operators category.
[18] Due to small sample sizes, only one co-morbid psychiatric illness can be represented in the model at a time.

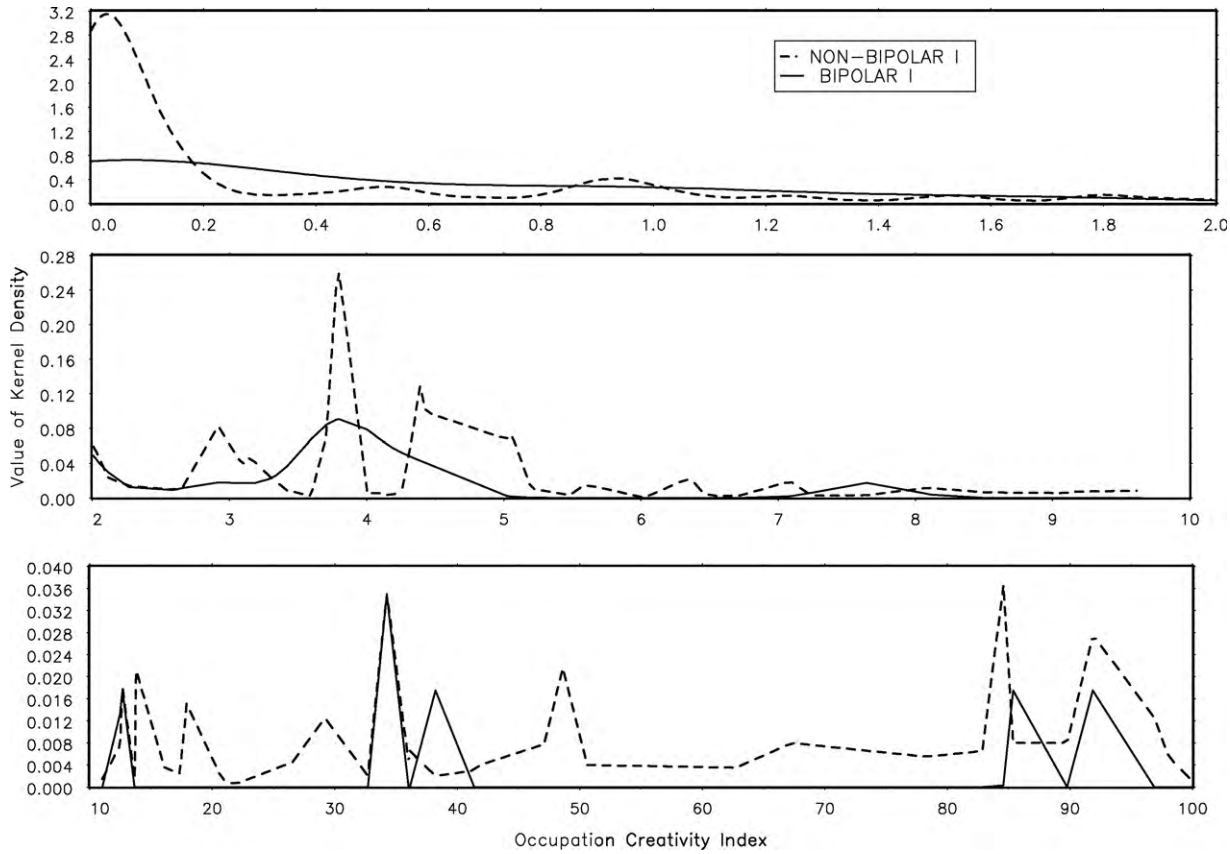


Fig. 1. Kernel density of the occupational creativity index with Silverman's bandwidth. *Note*: The value of 100 for the occupational creativity index indicates the highest level of creativity. The bandwidth selection method is from Silverman (1986).

marginal effects of bipolar illness do not change from the original base models.

The lower panel of Table 4 shows the results of the model that accounts for alcohol or drug abuse. Specifically, this model includes ALCDRGS, a dummy variable = 1 if the individual has a diagnosis of alcohol dependence and/or abuse, or drug dependence and/or abuse. The marginal effect estimates of manic episode status are similar in value as for the base model in the upper panel. Note that statistical significance on the marginal effect of the bipolar dummy is lost for the technical group (DSM-III-R) but gained for the operators group (DSM-III). For the 'creative' occupational categories, the estimate for the DSM-III bipolar dummy variable loses significance for services, but the estimate for the DSM-III-R bipolar variable gains significance at 10 percent for the professional and managerial group.

In summary, we have explored a number of alterations to the base model, and we find that the estimated marginal effects of bipolar illness are remarkably consistent across specifications. In particular, the influence of bipolar status on the probability of being in a professional and managerial occupation or a service occupation is uniformly positive. Next we turn to some additional evidence based on nonparametric tests and estimation.

3.3. Nonparametric evidence

Recall that we found large differences between the mean and median values of the occupational creativity index (Table 3) for bipolars and non-bipolars. Consequently, we turn to nonparametric tests and estimation techniques to address these distributional issues.

We first estimate Kolmogorov–Smirnov test statistics for the equality of occupational creativity distributions for bipolar and other individuals, and find corrected p-values of 0.103 for the DSM-III group and 0.519 for the DSM-II-R group. It is surprising that the p-values are not lower, based on our earlier results on the significant differences in the variances of the creativity index for bipolars and the rest of the sample for both DSM-III and DSM-III-R subsamples. We now use nonparametric kernel density estimation to further explore and test for differences in the occupational creativity index distribution for bipolars and other individuals.

3.3.1. Nonparametric density estimation

The advantage of nonparametric density estimation is that it is flexible and nonrestrictive. It can reveal the underlying density of a variable, even when that density deviates from the traditional normal form often assumed in parametric estimation. As the particular density of a

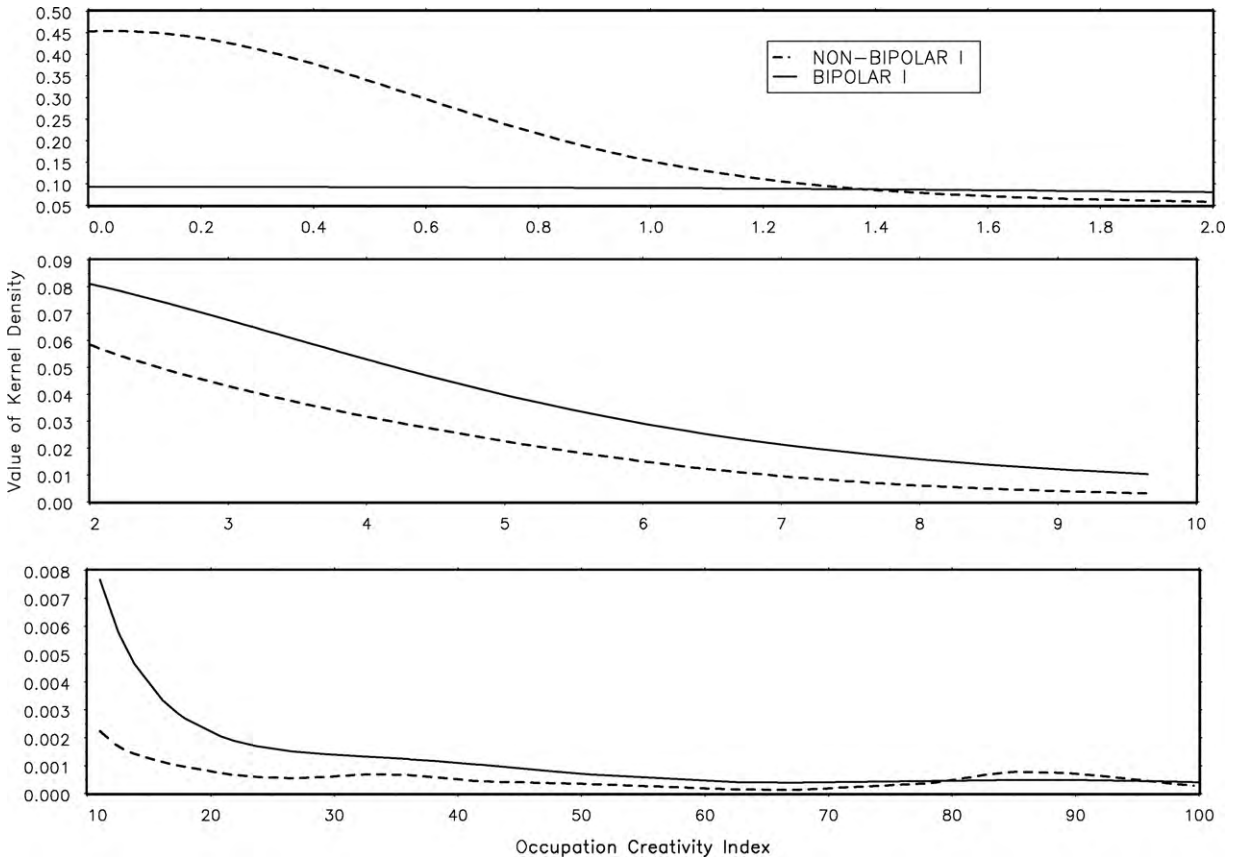


Fig. 2. Kernel density of the occupational creativity index with adaptive bandwidth.

variable need not be assumed a priori, specification errors can be avoided.

Here we estimate nonparametric kernel densities for the occupational creativity index for the bipolar and non-bipolar groups. For the kernel, we choose the Gaussian (standard normal) density function. The more difficult issue is the choice of bandwidth. We begin with the automatic bandwidth choice method described in Silverman (1986), which minimizes the integrated squared error (ISE) between the estimated function and the objective function. The kernel estimates using Silverman's plug-in bandwidth are displayed in Fig. 1.[19] The estimated densities are segmented into three intervals: [0,2) in the upper panel, [2,10) in the middle panel, and [10,100] in the lower panel. The upper panel highlights the heavy concentration of the data near zero, indicating that many workers do not perform creative activities in their jobs. The density is far less concentrated near zero for bipolar than for non-bipolar workers, however, evidence of greater occupational creativity for bipolars than non-bipolars. Note that the values of the kernel densities on the *y*-axis in the three panels dramatically decline from the upper to the middle to the lower panels. This reflects the skewness and lengthy tails of the distributions. Although Silverman's bandwidth produced a smooth density for values of the occupational creativity variable less than 2, the lower panels show that from 2 and beyond, the tails of both kernel densities are undersmoothed.

To focus on the tail distributions without oversmoothing the rest of the distribution, we employ another bandwidth selection method known as the 'adaptive two stage estimator' (A2SE), first considered by Breiman et al. (1977) and Abramson (1982). The basic idea is to use a broader bandwidth in the regions of low density to smooth the tail part of the function. Our results using this method are displayed in Fig. 2. Again we separate the kernel density into the intervals, [0,2), [2,10) and [10,100].[20] It is clear from the upper panel that the non-bipolar population is more concentrated at the lowest creativity levels. Over the 2–10 and 10–80 ranges of the creativity index, the bipolar population has higher density. By inspection, the

[19] Silverman's plug-in bandwidth is calculated as $h = 0.79RN^2$ where R is the inter-quartile range of the data and N is the sample size.

[20] The range of numbers on the *y*-axes in Fig. 1 differs from that in Fig. 2 because Silverman's bandwidth is much smaller than the A2SE bandwidth, on average. The area under a kernel density plot is one by design. With a larger bandwidth (A2SE), the weight of the density is allocated more evenly over the values of the creativity index in order to get a smoother curve, whereas with a smaller bandwidth (Silverman's) the weight is concentrated near the origin where most of the data points are located.

Table 5
Occupation creativity index dist. (c) by bipolar status.

Probability	Bipolar	Non-bipolar
$P(c \geq 1)$	0.3095	0.2599
$P(c \geq 2)$	0.1786	0.1627
$P(c \geq 3)$	0.1667	0.1424
$P(c \geq 4)$	0.1071	0.0922
$P(c \geq 5)$	0.0833	0.0724
$P(c \geq 6)$	0.0833	0.0576
$P(c \geq 7)$	0.0833	0.0531
$P(c \geq 8)$	0.0714	0.0495
$P(c \geq 7)$	0.0714	0.0466
$P(c \geq 8)$	0.0714	0.0441

Note: the highest value of c is 100.

area between the two densities from 10 to 80 (and about 95 to 100) exceeds the area from 80 to 95 where the bipolar density lies below the non-bipolar density. We also calculate the probabilities of observing the occupational creativity index above various values (integers 1 through 10) for each group. These are displayed in Table 5. The median level of the occupational creativity index (c) lies between 0.12 and 0.46 for the non-bipolar and bipolar samples, respectively (Table 3). The probability that c is greater than or equal to one in column (1) represents part of the "tail" above the median. We see that the probability of observing an occupational creativity index greater than one is higher for bipolars than non-bipolars (0.31 versus 0.26). The probability is higher for bipolars for every value of the creativity index up to greater than or equal to 10, which represents 4–7 percent of the samples.

To test if these densities are significantly different, we follow the method proposed by Li (1996). Let the two densities be $f(x)$ and $g(x)$, then test the null hypothesis $H_o : f(x) = g(x)$ against $H_1 : f(x) \neq g(x)$. The idea is to construct a nonparametric approximation for the integrated squared difference between the two estimated functions, i.e., the test statistic is based on $I = \int (\hat{f} - \hat{g})^2$. For our sample, the estimated test statistic is $Z = -3.01$ with a p-value of 0.001, so $H_o : f(x) = g(x)$, that the density functions of the occupational creativity index for non-bipolars and bipolars is the same, is rejected. That is, the densities are highly significantly different. This provides evidence of creativity differences for bipolar and non-bipolar persons.

4. Conclusion

Earlier work investigating links between occupational creativity and bipolar illness is largely based on psychiatric case studies or very small samples. In contrast we use a population based data set and econometric techniques to analyze occupational outcomes for those with bipolar disorder and the general population. We find evidence that bipolars are concentrated in service occupations as well as in professional and managerial occupations. We match the occupations to a measure of occupational creativity from the *Dictionary of Occupational Titles*. The professional and managerial occupations which include artists, musicians, and authors, are the most creative occupational category based on this index. We also find significantly higher median occupational creativity (DSM-III group) and education levels (both groups) for bipolars than for the general population.

After establishing the connection between occupation and creativity, we investigate the occupational outcomes of bipolars versus others in our sample using multinomial logit regressions. The results show that the effect of the bipolar dummy variable for a manic episode is positive in the most creative occupations: (1) professional and managerial, and (2) services. This result is robust to a number of alternative specifications.

We employ nonparametric density estimation methods to compare the densities of the occupational creativity index for the bipolar and non-bipolar samples. Based on these estimates we reject the null hypothesis that the densities are the same for bipolar I's and non-bipolars at the 0.1% level of significance. Probabilities based on these estimates also suggest that the data are more likely to reside in the tail of the creativity index density for bipolar than non-bipolar workers. That is, the probability of engaging in creative activities on the job is higher for bipolar than non-bipolar workers.

Still a number of caveats are in order. Some of the hypothesis tests here resulted in insignificant or marginally significant test statistics. Longitudinal data would better reveal the time patterns of illness onset, education, marital states and career paths. Selectivity issues, such as observing creativity of only employed or previously employed bipolar individuals, with longitudinal data as well. Future studies would also benefit from more recent data including the current diagnostic criteria (DSM-IV-TR).

Nevertheless we do find substantive evidence of a link between bipolar illness and occupational creativity using quite different data and methods than previous studies. This implies that productivity gains from enhanced creativity might outweigh productivity losses resulting from bipolar illness (for example, from absenteeism and lack of energy during depressive states). Thus, employers might be more willing to comply with the Americans with Disabilities Act of 1990, which prohibits discrimination against qualified persons on the basis of disability, including bipolar disorder, in all employment practices, and mandates employer provision of accommodations, such as modified work schedules and job reassignment, to enable performance of essential job functions.[21] Understanding their potential contributions might encourage employers to more readily hire and facilitate the productive capacities of individuals with bipolar illness.

The extent to which bipolar persons contribute to the stock of creative human capital in the workforce has implications for society's response to the inevitable identification and understanding of the genetic markers for bipolar disorder. The creative benefits of the disease is one factor weighing against its elimination via genetic

[21] U.S. Equal Employment Opportunity Commission (1997). Another example of disability employment policy is the U.K. Disability Discrimination Act (U.K. Government website, http://www.direct.gov.uk.) See Tremblay (2008) for accommodations that enhance performance of bipolar workers.

engineering or selective abortion (Goodwin and Jamison, 2007: 405–406, 461). The development of medication that reduces the social and private costs of the illness without inhibiting creativity might be a fruitful social investment.

Acknowledgements

The authors gratefully acknowledge helpful comments from Ronald Oaxaca, James F. Ragan, Jr., Victor J. Tremblay, James Walker, and our referees. We thank Adonis Yatchew for encouraging us not to give up on density estimation and Dr. James Phelps for providing medical expertise. The usual caveat applies.

References

American Psychiatric Association, 2000. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. The Association, Washington, D.C..

American Psychiatric Association, 1987. Diagnostic and Statistical Manual of Mental Disorders, Third edition, Revision. The Association, Washington, D.C..

American Psychiatric Association, 1980. Diagnostic and Statistical Manual of Mental Disorders, Third edition. The Association, Washington, D.C..

Abramson, I.S., 1982. On bandwidth variation in kernel estimates—a square root law. Annals of Statistics 10, 1217–1223.

Blau, F.D., Ferber, M.A., Winkler, A.E., 1998. The Economics of Women, Men and Work, 3rd edition. Prentice Hall, Upper Saddle River, NJ.

Breiman, L., Weisel, W., Purcell, E., 1977. Variable kernel estimates of multivariate densities. Technometrics 19, 135–144.

England, P., Kilbourne, B., 1988. Occupational Measures from the Dictionary of Occupational Titles for 1980 Census Detailed Occupations [computer file]. Inter-university Consortium for Political and Social Research [distributor], Ann Arbor, MI.

Ettner, S.L., Frank, R.G., Kessler, R.C., 1997. The impact of psychiatric disorders on labor market outcomes. Industrial and Labor Relations Review 51 (1), 64–81.

Filer, R.K., 1986. The 'starving artist'—myth or reality? Earnings of artists in the United States. Journal of Political Economy 94 (1), 56–75.

Frank, R., McGuire, T., 2000. Economics and mental health. In: Culyer, A.J., Newhouse, J.P. (Eds.), Handbook of Health Economics, vol. 1B. Elsevier, Holland.

Gottfredson, G.D., Holland, J.L., 1996. Dictionary of Holland Occupational Codes. Psychological Assessment Resources, Lutz, Florida.

Goodwin, F.K., Jamison, K.R., 2007. Manic-Depressive Illness, 2nd edition. Oxford University Press, New York.

Jamison, K.R., 1993. Touched with Fire: Manic-Depressive Illness and the Artistic Temperament. Simon and Schuster, New York.

Judd, L.L., Akiskal, H.S., 2003. The prevalence and disability of bipolar spectrum disorders in the US population: re-analysis of the ECA database taking into account subthreshold cases. Journal of Affective Disorders. 73, 123–131.

Keller, M., Miller, G.F., 2006. Which evolutionary genetic models best explain the persistence of common, harmful, heritable mental disorders? Behavioral and Brain Sciences 29, 385–404.

Kèri, Szaboles, 2009. Genes for psychosis and creativity. Psychological Science 20 (9), 1070–1073.

Li, Q., 1996. Nonparametric testing of closeness between two unknown distribution functions. Econometric Reviews 5, 261–274.

Marcotte, D.E., Wilcox-Gök, V., Redmon, D.P., 2000. The labor market effects of mental illness: the case of affective disorders. In: Sorkin, A. (Ed.), Research in Human Capital and Development, vol. 13. JAI Press Inc., Stamford, Conn..

Merikangas, K.R., Akiskal, H.S., Angst, J., 2007. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. Archives of General Psychiatry 64 (5), 543–552.

Miller, G.F., Tal, I., 2007. Schizotypy versus intelligence and openness as predictors of creativity. Schizophrenia Research 93 (1–3), 317–324.

Richards, R., Kinney, D.K., 1988. Creativity in manic-depressives, cyclothymes, their normal relatives and control subjects. Journal of Abnormal Psychology 97 (3), 281–288.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, New York.

Tremblay, C.H., 2008. Workplace accommodations and job success for persons with bipolar disorder. Working paper, Oregon State University.

U.K. Government website, 2010. Employment Rights and the Disability Discrimination Act. Website: http://www.direct.gov.uk, Crown copyright.

U.S. Department of Health and Human Services, National Institute of Mental Health. Epidemiologic Catchment Area Study, 1980–1985 [computer file], Rockville, MD: NIMH [producer] Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1991, Persistent URL: http://dx.doi.org/10.3886/ICPSR08993.

U.S. Department of Labor, Manpower Administration, 1972. Handbook for Analyzing Jobs.

U.S. Equal Employment Opportunity Commission, EEOC enforcement guidance on the Americans with Disabilities Act and psychiatric disabilities. Notice Number 915.002, March 25, 1997.

Weissman, M.M., Bruce, M.L., Leaf, P.J., 1991. Affective disorders. In: Robins, L.N., Regier, D.A. (Eds.), Psychiatric Disorders in America. Free Press, New York.

Zimmerman, M., 2008. Is bipolar disorder overdiagnosed? Journal of Clinical Psychiatry 69, 935–940.

Please cite this article in press as: Sandrini et al., Causal Role of Prefrontal Cortex in Strengthening of Episodic Memories through Reconsolidation, Current Biology (2013), http://dx.doi.org/10.1016/j.cub.2013.08.045

Current Biology 23, 1–4, November 4, 2013 ©2013 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2013.08.045

Report

Causal Role of Prefrontal Cortex in Strengthening of Episodic Memories through Reconsolidation

Marco Sandrini,[1,2,*] Nitzan Censor,[1] Jonathan Mishoe,[1] and Leonardo G. Cohen[1]

[1]Human Cortical Physiology and Stroke Neurorehabilitation Section, National Institute of Neurological Disorders and Stroke (NINDS), National Institutes of Health (NIH), Bethesda, MD 20892, USA

[2]Center for Neuroscience and Regenerative Medicine, Uniformed Services University of Health Sciences, Bethesda, MD 20814, USA

Summary

Memory consolidation is a dynamic process. Reactivation of consolidated memories triggers reconsolidation, a time-limited period during which memories can be modified [1–4]. Episodic memory refers to our ability to recall specific past events about what happened, including where and when [5]. However, it is unknown whether noninvasive stimulation of the neocortex during reconsolidation might strengthen existing episodic memories in humans. To modify these memories, we applied repetitive transcranial magnetic stimulation (rTMS) [6] over right lateral prefrontal cortex (PFC), a region involved in the reactivation of episodic memories [7, 8]. We report that rTMS of PFC after memory reactivation strengthened verbal episodic memories, an effect documented by improved recall 24 hr postreactivation compared to stimulation of PFC without reactivation and vertex (control site) after reactivation. In contrast, there was no effect of stimulation 1 hr postreactivation (control experiment), showing that memory strengthening is time dependent, consistent with the reconsolidation theory. Thus, we demonstrated that right lateral PFC plays a causal role in strengthening of episodic memories through reconsolidation in humans. Reconsolidation may serve as an opportunity to modify existing memories with noninvasive stimulation of a critical brain region, an issue of fundamental importance for memory research and clinical applications.

Results and Discussion

Consolidation theory assumes that memories are unstable (i.e., susceptible to interference) for a limited time after encoding, but as time passes, memories stabilize and become resistant to interference [9]. Newly formed memories require gene expression for several hours in order to become stable or consolidated [3]. However, studies done in the last decade demonstrated that the initial gene-expression-dependent phase is not the only one [1–3]. In fact, consolidated memories can re-enter unstable states when they are reactivated during retrieval or by a reminder cue and need to consolidate again in order to persist over longer periods of time [1–4]. Thus, the concept of reconsolidation assumes that memories are not consolidated once and forever, challenging the view that stability characterizes consolidated memories [2–4]. Indeed, memory reactivation triggers reconsolidation, a time-limited period during which existing memories are susceptible to modifications [1–4]. Thus, memories can be stabilized, strengthened [10–13], weakened/disrupted [1, 14, 15], or updated by the inclusion of new information [16, 17] through reconsolidation.

Episodic memory refers to our ability to recall specific details about past events, including where, when, and what happened [5]. The prefrontal cortex (PFC) is a critical node in the neural network mediating the encoding and retrieval of these memories [7, 18].

Although previous human studies on reconsolidation have shown that existing episodic memories can be enhanced by stressor events [10], emotional processing [11], successive reactivations of the memories [12], or pharmacological modulations [13], it remains unknown whether the lateral PFC plays a causal role in the reconsolidation process.

To modify episodic memories, we applied repetitive transcranial magnetic stimulation (rTMS) over right lateral PFC. Since retrieval or reactivation by a reminder cue triggers reconsolidation, we decided to stimulate a neocortical region involved in retrieval [7, 18] and reactivation of episodic memories [8]. Previous rTMS studies demonstrated the causal role of right, but not left, lateral PFC in retrieval [7, 18], and these findings were confirmed in a functional resonance magnetic imaging (fMRI) study showing activation of the right, but not left, lateral PFC during memory reactivation by a reminder cue [8]. rTMS is a noninvasive brain stimulation technique used in this protocol to evaluate the causal role of focal neocortical regions and to modulate cognitive function [6, 19] (see the Supplemental Experimental Procedures available online).

Subjects (n = 30) learned a list of 20 object words on day 1. This procedure was repeated until the participants remembered at least 17 of the 20 words (85%) or until a maximum of four learning trials was reached [16, 17]. On day 2, in a subset of subjects (n = 10), existing memories were reactivated by a spatial-contextual reminder cue (without explicit recall), and 10 min later 1 Hz rTMS was applied for 15 min to right dorsolateral PFC (PFC-R) (see the Supplemental Experimental Procedures). There is evidence that memories are automatically reactivated if the original learning spatial context is part of the reminder [17]. To determine whether the rTMS effect was specific to memory reactivation and relied on right PFC function, we designed two control conditions that were applied to the remaining subjects (n = 10 per group). First, to determine whether the rTMS effect was specific to memory reactivation, we applied rTMS over right PFC without memory reactivation (PFC-NR), a behavioral manipulation previously successfully done in animal [1] and human [16] reconsolidation studies. Second, to determine whether the rTMS effect was topographically specific, we applied rTMS over the vertex after memory reactivation (vertex-R, control site) [6]. We chose the frequency of rTMS on the basis of a recent study showing that 1 Hz rTMS of right lateral PFC enhanced episodic memory performance when delivered during the interval (10 min) between encoding and retrieval [20]. Memory recall was tested on Day 3 (24 hr postreactivation) (see Figure 1 and the Supplemental Experimental Procedures).

*Correspondence: marco.sandrini@nih.gov

Please cite this article in press as: Sandrini et al., Causal Role of Prefrontal Cortex in Strengthening of Episodic Memories through Reconsolidation, Current Biology (2013), http://dx.doi.org/10.1016/j.cub.2013.08.045

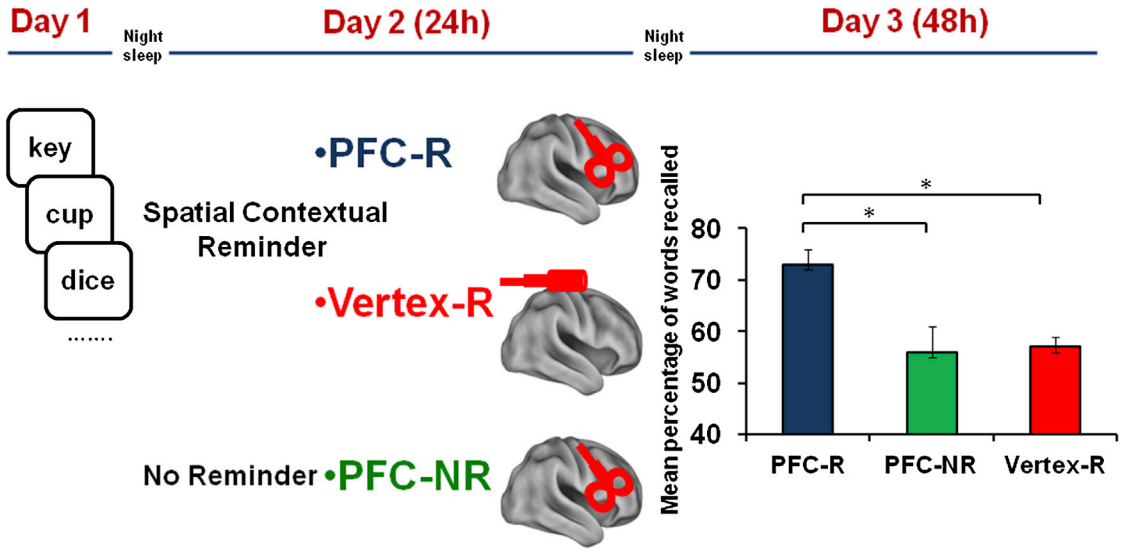


Figure 1. Timeline and Results of Experiment 1

Subjects learned 20 words on day 1. On day 2, rTMS was applied to right dorsolateral PFC (PFC-R) or a control site (vertex-R) after memory reactivation by a spatial-contextual reminder. A third group of subjects received rTMS of right PFC without memory reactivation (PFC-NR). Recall was tested on day 3. The plot shows the mean percentage of words recalled in each group (n = 10). Error bars represent the SEM. *p < 0.0167. Cortical surfaces were generated using CARET (http://brainvis.wustl.edu/wiki/index.php/Main_Page).

In order to compare the learning rate of the three experimental groups, we recorded how many learning trials (1–4) were necessary for participants to recall at least 17 words (85%) on day 1. Participants needed on average 3.4 learning trials to reach this criterion (PFC-R = 3.3 ± 0.26 [mean ± SEM]; PFC-NR = 3.5 ± 0.22; vertex-R = 3.4 ± 0.27). There were no significant differences between the three groups (H = 0.133, p = 0.93).

Regarding the number of words recalled on day 3 (see the Supplemental Experimental Procedures), the main effect "group" was significant [F(2,27) = 6.30, p < 0.01]. Independent t tests (Bonferroni corrected, p = 0.0167) showed significant differences between PFC-R (14.3 ± 0.63) and PFC-NR (11.3 ± 1.05) (p < 0.0161) and between PFC-R and vertex-R (11.32 ± 0.45) (p < 0. 0006). No difference was found between PFC-NR and vertex-R (p = 0.9658). To enable comparison between experiments, we have displayed the mean percentage of words recalled in each group in Figure 1 (PFC-R = 73% ± 3.17%; PFC-NR = 56.3% ± 5.24%; vertex-R = 56.6% ± 2.27%).

These findings show that stimulation of right PFC after memory reactivation (PFC-R) strengthened verbal episodic memories, an effect indicated by enhanced later recall (24 hr postreactivation) compared to rTMS of PFC without memory reactivation (PFC-NR) or rTMS of a control site after memory reactivation (vertex-R).

To determine whether the rTMS effect (PFC-R) was reconsolidation specific, we carried out a control experiment (experiment 2). Because reconsolidation is a process requiring gene expression for several hours [2, 3], memory modification should not be expressed shortly postreactivation (i.e., 1 hr, when the memories are still unstable) but many hours later (i.e., 24 hr, after memories have become stable again) [2, 3, 13, 16]. In animal studies, memory reconsolidation is only observed after a delay of about 4 hr [2].

To address this question, in experiment 2 we used exactly the same materials and followed exactly the same procedure as in experiment 1 with one exception: the experiment consisted of only two experimental sessions in consecutive days (24 hr apart). The procedure described for day 3 in experiment 1 (i.e., memory recall) was administered on day 2 shortly (i.e., 30 min) after the end of the rTMS session (1 hr postreactivation). This 30 min interval was chosen to allow the aftereffects of rTMS to wash out in order to avoid effects on the retrieval process per se [6]. Based on the findings of experiment 1, we compared right PFC-R (n = 10) to vertex-R (n = 10) (see Figure 2).

Subjects (n = 20) needed on average three learning trials to recall at least 17 words (85%) (PFC-R = 3 ± 0.30; vertex-R = 3 ± 0.26). No significant differences were found between the groups (U = 48.0, p = 0.87).

Regarding the number of words recalled on day 2 (see the Supplemental Experimental Procedures), the main effect of "group" [F(1,18) = 0.133, p = 0.72; PFC = 13.85 ± 0.95; vertex = 13.4 ± 0.95] was not significant. The mean percentage of words recalled in each group is displayed in Figure 2 (vertex-R = 67% ± 3.94%; PFC-R = 69.25% ± 4.74%).

Figure 2 shows that, in opposition to findings with recall on day 3 (24 hr postreactivation), testing shortly (1 hr) postreactivation on day 2 yielded comparable results for both vertex-R and PFC-R stimulation. These results suggest that the effect of experiment 1 is time dependent because it is only seen if the memory trace is allowed to reconsolidate.

In this study, we show that rTMS of right lateral PFC after memory reactivation (PFC-R) strengthens existing verbal episodic memories in humans, an effect indicated by enhanced later recall (24 hr postreactivation) compared to control groups (i.e., vertex-R and PFC-NR). Moreover, this effect does not occur when memory performance was tested shortly postreactivation (1 hr). Thus, as predicted by reconsolidation theory [2, 3], the effect of rTMS over right lateral PFC evolves over time and depends on memory reconsolidation.

In this study, we demonstrate for the first time that reconsolidation may serve as an opportunity to strengthen existing

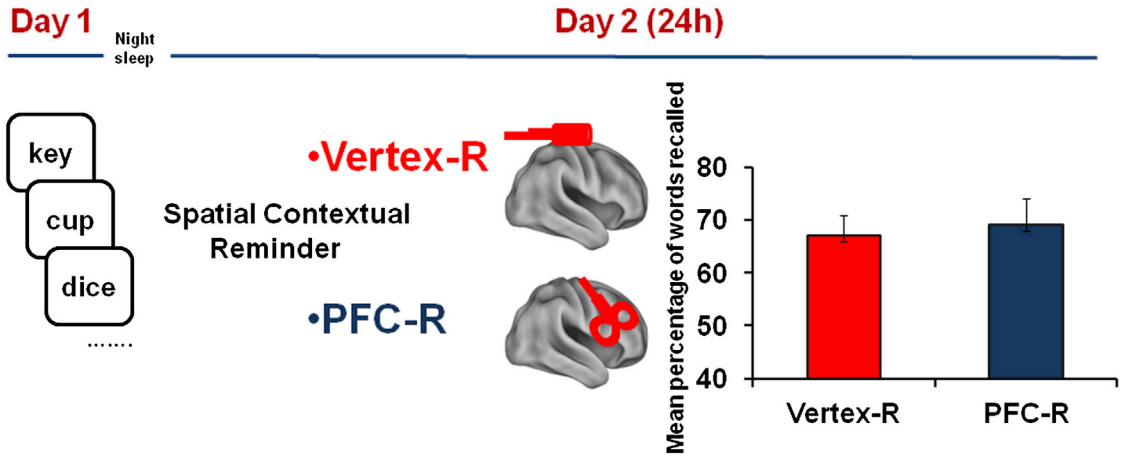


Figure 2. Timeline and Results of Experiment 2

Subjects learned 20 words on day 1. On day 2, rTMS was applied to right dorsolateral PFC (PFC-R) or a control site (vertex-R) after memory reactivation by a spatial-contextual reminder. Recall was tested shortly (30 min) after rTMS. The plot shows the mean percentage of words recalled in each group (n = 10). Error bars represent the SEM. Cortical surfaces were generated using CARET (http://brainvis.wustl.edu/wiki/index.php/Main_Page).

episodic memories with non-invasive stimulation of the right lateral PFC, showing that enhanced recollection can be achieved by targeting a critical brain region during a postreactivation state of plasticity.

Regarding the putative brain mechanisms of this effect, the role played by spatial context in memory reactivation [17] points to hippocampal involvement in this paradigm [4] and suggests possible functional interactions between this brain region and PFC during the reconsolidation process [21]. It is well recognized that memory consolidation involves a relatively brief cascade of molecular and cellular events that alter synaptic efficacy as well as a prolonged systems level interaction between the hippocampus and neocortex [4, 21]. Episodic memories initially consist of two components: content representations dependent on neocortical networks and a spatial-contextual representation dependent on the hippocampus [4], which seems to serve as the memory trace that permits the retrieval of memory content stored elsewhere [22].

Since rTMS affects not only the targeted local region but also activity in remote interconnected regions [6, 19], 1 Hz rTMS of right PFC after memory reactivation may have enhanced the functional coupling between the PFC and the hippocampus, thereby enhancing memory recollection through reconsolidation. There is evidence that 1 Hz rTMS may improve performance of a cognitive task by strengthening the connectivity between task-relevant brain regions depending on the functional state of the cortex at the time of stimulation [23]. Thus, combined TMS-fMRI studies may shed light on how functional interactions between remote but interconnected brain regions may support reconsolidation of episodic memories [6, 19]. Future work might identify the time window of memory reconsolidation by varying the interval between memory reactivation and rTMS and address the question of whether the same or different lateral PFC regions (i.e., left versus right and dorsal versus ventral) operate in both consolidation and reconsolidation.

We conclude that right lateral PFC plays a critical role in the neural network that mediates the strengthening of episodic memories through reconsolidation in humans.

Noninvasive stimulation of a critical brain region during reconsolidation may be a new opportunity for designing well-controlled, double-blinded interventional studies, which can potentially strengthen adaptive memories in elderly adults and individuals with memory dysfunctions (e.g., mild cognitive impairment) or disrupt maladaptive memories in individuals with posttraumatic stress disorder.

Experimental Procedures

The experimental procedures are summarized briefly throughout the Results and are presented in complete detail in the Supplemental Experimental Procedures.

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2013.08.045.

Acknowledgments

This work was supported by the Intramural Research Program of the NINDS, NIH to J.M. and L.G.C., by an NINDS Ruth L. Kirschstein National Research Service Award (NRSA) to N.C., and by funding from Department of Defense in the Center for Neuroscience and Regenerative Medicine (CNRM) to M.S. We would like to thank Sook-Lei Liew for insightful comments and suggestions.

Received: June 3, 2013
Revised: August 4, 2013
Accepted: August 19, 2013
Published: October 24, 2013

References

1. Nader, K., Schafe, G.E., and Le Doux, J.E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. Nature *406*, 722–726.
2. Nader, K., and Hardt, O. (2009). A single standard for memory: the case for reconsolidation. Nat. Rev. Neurosci. *10*, 224–234.
3. Alberini, C.M. (2011). The role of reconsolidation and the dynamic process of long-term memory formation and storage. Front Behav Neurosci *5*, 12.
4. Nadel, L., and Hardt, O. (2011). Update on memory systems and processes. Neuropsychopharmacology *36*, 251–273.
5. Tulving, E. (2002). Episodic memory: from mind to brain. Annu. Rev. Psychol. *53*, 1–25.

Please cite this article in press as: Sandrini et al., Causal Role of Prefrontal Cortex in Strengthening of Episodic Memories through Reconsolidation, Current Biology (2013), http://dx.doi.org/10.1016/j.cub.2013.08.045

6. Sandrini, M., Umiltà, C., and Rusconi, E. (2011). The use of transcranial magnetic stimulation in cognitive neuroscience: a new synthesis of methodological issues. Neurosci. Biobehav. Rev. *35*, 516–536.
7. Sandrini, M., Cappa, S.F., Rossi, S., Rossini, P.M., and Miniussi, C. (2003). The role of prefrontal cortex in verbal episodic memory: rTMS evidence. J. Cogn. Neurosci. *15*, 855–861.
8. Diekelmann, S., Büchel, C., Born, J., and Rasch, B. (2011). Labile or stable: opposing consequences for memory when reactivated during waking and sleep. Nat. Neurosci. *14*, 381–386.
9. McGaugh, J.L. (2000). Memory—a century of consolidation. Science *287*, 248–251.
10. Coccoz, V., Maldonado, H., and Delorenzi, A. (2011). The enhancement of reconsolidation with a naturalistic mild stressor improves the expression of a declarative memory in humans. Neuroscience *185*, 61–72.
11. Finn, B., and Roediger, H.L., 3rd. (2011). Enhancing retention through reconsolidation: negative emotional arousal following retrieval enhances later recall. Psychol. Sci. *22*, 781–786.
12. Forcato, C., Rodríguez, M.L., and Pedreira, M.E. (2011). Repeated labilization-reconsolidation processes strengthen declarative memory in humans. PLoS ONE *6*, e23305.
13. Rodríguez, M.L., Campos, J., Forcato, C., Leiguarda, R., Maldonado, H., Molina, V.A., and Pedreira, M.E. (2013). Enhancing a declarative memory in humans: the effect of clonazepam on reconsolidation. Neuropharmacology *64*, 432–442.
14. Chan, J.C., and LaPaglia, J.A. (2013). Impairing existing declarative memory in humans by disrupting reconsolidation. Proc. Natl. Acad. Sci. USA *110*, 9309–9313.
15. Censor, N., Dimyan, M.A., and Cohen, L.G. (2010). Modification of existing human motor memories is enabled by primary cortical processing during memory reactivation. Curr. Biol. *20*, 1545–1549.
16. Hupbach, A., Gomez, R., Hardt, O., and Nadel, L. (2007). Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. Learn. Mem. *14*, 47–53.
17. Hupbach, A., Hardt, O., Gomez, R., and Nadel, L. (2008). The dynamics of memory: context-dependent updating. Learn. Mem. *15*, 574–579.
18. Manenti, R., Cotelli, M., Robertson, I.H., and Miniussi, C. (2012). Transcranial brain stimulation studies of episodic memory in young adults, elderly adults and individuals with memory dysfunction: a review. Brain Stimulat. *5*, 103–109.
19. Dayan, E., Censor, N., Buch, E.R., Sandrini, M., and Cohen, L.G. (2013). Noninvasive brain stimulation: from physiology to network dynamics and back. Nat. Neurosci. *16*, 838–844.
20. Turriziani, P., Smirni, D., Zappalà, G., Mangano, G.R., Oliveri, M., and Cipolotti, L. (2012). Enhancing memory performance with rTMS in healthy subjects and individuals with Mild Cognitive Impairment: the role of the right dorsolateral prefrontal cortex. Front Hum Neurosci *6*, 62.
21. Wang, S.H., and Morris, R.G. (2010). Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. Annu. Rev. Psychol. *61*, 49–79, C1–C4.
22. Hardt, O., Wang, S.H., and Nader, K. (2009). Storage or retrieval deficit: the yin and yang of amnesia. Learn. Mem. *16*, 224–230.
23. Ward, N.S., Bestmann, S., Hartwigsen, G., Weiss, M.M., Christensen, L.O., Frackowiak, R.S., Rothwell, J.C., and Siebner, H.R. (2010). Low-frequency transcranial magnetic stimulation over left dorsal premotor cortex improves the dynamic control of visuospatially cued actions. J. Neurosci. *30*, 9216–9223.

ARTICLE OPEN

Check for updates

The default network is causally linked to creative thinking

Ben Shofty[1,8]✉, Tal Gonen[1,2,8], Eyal Bergmann[3], Naama Mayseless[4,5], Akiva Korn[1], Simone Shamay-Tsoory[4], Rachel Grossman[1], Itamar Jalon[2,6], Itamar Kahn[3,7,9] and Zvi Ram[1,9]

© The Author(s) 2021

Creative thinking represents a major evolutionary mechanism that greatly contributed to the rapid advancement of the human species. The ability to produce novel and useful ideas, or original thinking, is thought to correlate well with unexpected, synchronous activation of several large-scale, dispersed cortical networks, such as the default network (DN). Despite a vast amount of correlative evidence, a causal link between default network and creativity has yet to be demonstrated. Surgeries for resection of brain tumors that lie in proximity to speech related areas are performed while the patient is awake to map the exposed cortical surface for language functions. Such operations provide a unique opportunity to explore human behavior while disrupting a focal cortical area via focal electrical stimulation. We used a novel paradigm of individualized direct cortical stimulation to examine the association between creative thinking and the DN. Preoperative resting-state fMRI was used to map the DN in individual patients. A cortical area identified as a DN node (study) or outside the DN (controls) was stimulated while the participants performed an alternate-uses-task (AUT). This task measures divergent thinking through the number and originality of different uses provided for an everyday object. Baseline AUT performance in the operating room was positively correlated with DN integrity. Direct cortical stimulation at the DN node resulted in decreased ability to produce alternate uses, but not in the originality of uses produced. Stimulation in areas that when used as network seed regions produced a network similar to the canonical DN was associated with reduction of creative fluency. Stimulation of areas that did not produce a default-like network (controls) did not alter creative thinking. This is the first study to causally link the DN and creative thinking.

Molecular Psychiatry; https://doi.org/10.1038/s41380-021-01403-8

INTRODUCTION

Creativity, traditionally described as the ability to produce novel and useful ideas [1], has played a major role in the establishment of modern culture and civilization. Creative thinking is essential to the daily functioning of modern human beings and allows for inspirational problem solving, technological and artistic advancement. This complex, multi-factorial cognitive process cannot be linked to a specific cerebral anatomical area, such as primary motor or sensory cortices, or to the activity of individual neurons, and instead is assumed to be network-dependent [2]. With the development of robust whole-brain functional imaging, higher cognitive functions are increasingly associated with large-scale cortical networks. These networks are comprised of distributed neuronal ensembles that are spatially dispersed but temporally synchronized [3]. Large-scale neuronal networks are nowadays decipherable using advanced functional imaging and computational methods and are thought to subserve complex abilities, such as attention, executive function, and more [4–6]. The Default Network (DN) is a well-explored, large-scale, association cortical network traditionally linked with internal mentation [7, 8]. Its increased activity is associated with decreased performance in externally driven tasks, and it conversely demonstrates task-induced deactivation when focus is shifted to external stimuli [9]. Recent evidence suggests that connectivity between areas traditionally associated with the DN may underlie the ability to carry out creative thinking [1, 10]. In addition, DN interactions with the salience and executive networks, ensembles that are not traditionally functionally linked, are thought to underlie creative thinking [11].

Heretofore, despite many studies demonstrating correlation between the DN, other large-scale cortical networks, and creativity, no causal relation has been shown. While there have been publications describing transcranial disruption of an anatomical area, and associated impairment in DN related functions such as episodic memory [12] and simulation [13], these have targeted the canonical DN and did not use individualized DN mapping. In addition, several technical limitations associated with transcranial stimulation limit the ability to establish such causality or identify relationships at the individual participant level.

We utilized the unique experimental platform of awake brain surgery for resection of a primary brain tumor to explore the effects of DN node stimulation. During routine surgery near

[1]Department of Neurosurgery, Tel Aviv Sourasky Medical Center and The Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. [2]Functional Brain Center, Wohl Institute for Advanced Imaging, Tel Aviv Sourasky Medical Center, and The Sagol Center for Neuroscience, Tel Aviv University, Tel Aviv, Israel. [3]Department of Neuroscience, Rappaport Faculty of Medicine and Institute, Technion—Israel Institute of Technology, Haifa, Israel. [4]Department of Psychology, University of Haifa, Haifa, Israel. [5]School of Medicine, Stanford University, Stanford, CA, USA. [6]Department of Psychology, Tel Aviv University, Tel Aviv, Israel. [7]Present address: Department of Neuroscience, Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, USA. [8]These authors contributed equally: Ben Shofty, Tal Gonen. [9]These authors jointly supervised this work: Itamar Kahn, Zvi Ram. ✉email: Shoftyben@gmail.com

Received: 16 May 2021 Revised: 14 November 2021 Accepted: 24 November 2021
Published online: 01 January 2022

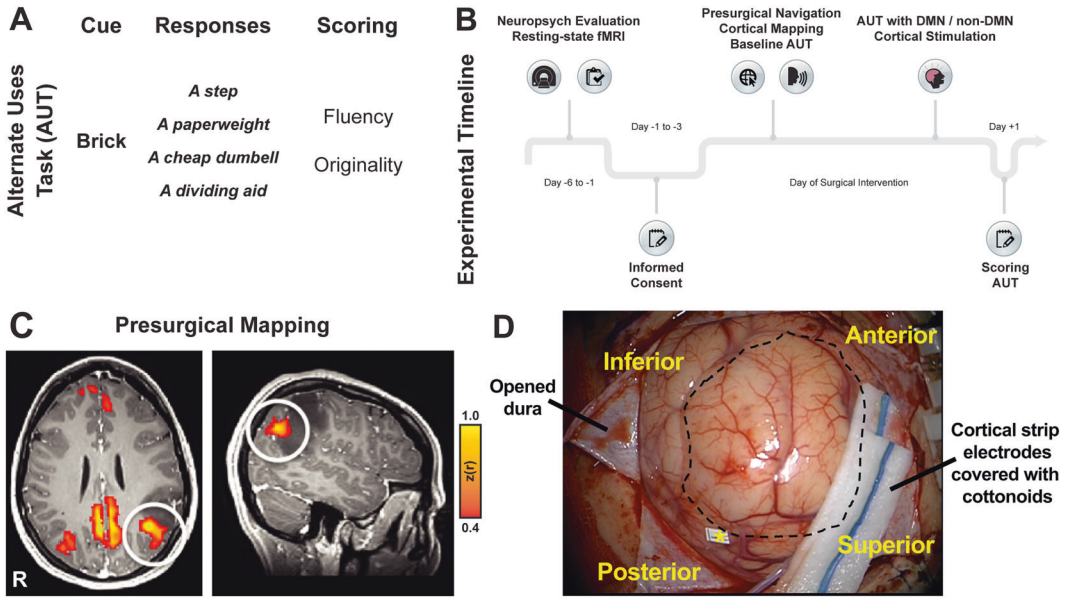


Fig. 1 Study design. A Alternate uses task (AUT). An example of a task cue, followed by potential common alternate uses a participant may produce. Each item was scored for fluency and originality creativity sub-components. **B** The experimental timeline from participant recruitment to postoperative period. **C** Individualized default network (DN) map. Seed-based whole-brain correlation maps were derived from the posterior cingulate cortex (PCC) region of interest. For each study participant, similar DN functional connectivity was derived individually and a target stimulation site was classified either as DN node (white circle) or non-DN site (i.e., outside the correlation map; not shown here). Stimulation sites were chosen such that they would be accessible intraoperatively. **D** An example intraoperative view of exposed tumor (dotted black line represents tumor surface circumference) and the target DN node for direct cortical stimulation as identified by intraoperative navigation (yellow asterisk).

eloquent brain regions, cortical mapping via electrical stimulation (direct cortical stimulation, DCS) is performed to identify functional cortical areas and thus preserve crucial functions, such as speech. Traditionally, intraoperative mapping identifies basic functions of sensation, motor, and speech; higher cognitive functions that are network-dependent are not mapped. DCS allows for the temporary inhibition of a limited cortical area using a bipolar stimulator that minimizes current spread to a ~2 mm sphere [14]. To investigate the possible connection between the DN and creative thinking, we examined the effect of DCS of a preoperatively-identified DN node on the ability to perform creative thinking. Patients who agreed to participate in this study underwent DCS of an individually mapped DN node (study) as extracted from preoperative resting state-based functional connectivity fMRI mapping of this network, or a non-DN stimulation (control). To evaluate creative thinking, we utilized the *alternate uses task* (AUT; Fig. 1A), a well-validated creativity task in which the subject needs to provide varied, alternate possible uses for everyday common objects [15]. The AUT assesses divergent thinking, which is widely considered to be an important antecedent of creativity because it involves the ability to consciously generate new ideas that branch out and allow for many possible solutions to a given problem, and is scored based on the originality and the number of valid uses [15]. The effects of stimulation on two main components of creative thinking, originality and creative fluency, as expressed by the relative domains in the AUT, were examined. We hypothesized that disruption of DN synchronization by artificial non-physiologic electrical stimulation would affect the ability to perform creative thinking (Fig. 1B).

METHODS

Patient selection and experimental procedure

All patients participating in this experiment willfully signed an informed consent form, and this experiment was approved by the Tel Aviv Medical Center ethics committee (0378-16-TLV). Included were patients with gliomas that were located in or near eloquent brain areas in the dominant hemisphere and were scheduled for clinically-indicated awake craniotomy for resection of their tumor. All patients had intact language functions preoperatively, and Hebrew was their native tongue. All patients underwent a thorough pre-operative neuropsychological evaluation, fMRI for determination of language dominance, and mapping of their DN (see below for details). Neurocognitive profiles were assessed preoperatively using the NeuroTrax computerized cognitive testing battery (NeuroTrax Corp., Bellaire, TX) [16]. Each patient's DN map was uploaded to the intraoperative neuro-navigation system (BrainLab AG, Feldkirchen, Germany) and used clinically to locate the tumor and plan the craniotomy and subsequent resection of the tumor. After rigid positioning in a head holder in the operating room, each patient underwent a baseline evaluation of language functions. Next, a tumor-tailored craniotomy was done under regional nerve blocks and local anesthetics as is customary in our center [17, 18]. Mapping for language functions was carried out following craniotomy and dural opening, covering the entire exposed cortical surface. Intraoperative mapping was carried out in two stages. First, in order to control for interference with language functions, naming, verb generation, and comprehension were tested. Language function baselines were established inside the operating room following cortical exposure, but before cortical stimulation, to control for anxiety and drug related effects on performance and difference in experimental setting. If any language effects were noted during mapping, or if clinical/sub-clinical electrical seizures were identified (by intraoperative electrocortiography) during DCS the participant was excluded from this study. Then, a baseline AUT consisting of five items was performed with no stimulation. Next, we located the DN node as mapped preoperatively using the intraoperative navigation system and, when possible, stimulated the node while the participant performed the AUT. The final intraoperative location of the DCS was documented on the neuro-navigation system. For control participants who did not have a DN node close to the craniotomy, we stimulated a cortical area that was verified to be outside the DN as mapped preoperatively, and the stimulation site was documented. Each participant's performance on the AUT was subsequently scored by a neuropsychologist (blind to the stimulation location) and results were compared to baseline and between study and control groups.

DN mapping

Brain imaging was performed at the Wohl Institute for Advanced Imaging, Tel Aviv Medical Center, using a Siemens MAGNETOM Prisma 3.0 T scanner

with a twenty-channel head coil. Anatomical 3D T1-weighted imaging was obtained using SPGR/FLASH sequences with 1 mm isotropic voxels. All functional whole-brain scans were performed with gradient echo-planar imaging (EPI) sequence of functional T2*-weighted images.

Scan parameters, preprocessing and functional connectivity analysis

For each participant, a six-minute resting-state fMRI scan was acquired (TR = 3000 ms, TE = 35 ms, flip angle = 90°, 44 slices at 3 mm thickness with no gap, 117 repetitions). Preprocessing was done in agreement with widely available pipelines, and as previously described [19, 20]. A functional connectivity measure of the DN integrity was computed as the correlation between time courses in two DN ROIs, posterior cingulate cortex (PCC) and medial prefrontal cortex (mPFC). Specifically, spherical ROIs with 6 mm radius were defined with centers at: $x = 0$, $y = -60$, $z = 30$ for PCC and −1, 49, 5 for mPFC (coordinates defined in the Montreal Neurological Imaging [MNI] atlas space). Individual DN integrity was calculated as the whole brain connectivity of the PCC ROI (thresholded at z (r) of 0.4). Similarity between individual DN maps derived from a stimulation ROI and the canonical DN was estimated using the Sørensen-Dice similarity coefficient between the stimulation seed-based map (thresholded at $z(r)$ of 0.2) and the DN from Yeo et al. [3].

Alternate uses task (AUT)

Different items for the AUT were used at baseline and during stimulation to avoid learning effects and were identical across all participants (baseline—shoe, button, bedsheet, nail, paper clip; stimulation—box, car tire, cup, pencil, can). Participants were blinded to the study rationale and task and were informed that the study goals were to improve the ability to avoid resection-induced deficits on language and related high-order cognitive functions. Patients were only instructed to provide alternate uses for the sample item (newspaper, and its possible alternate uses – start a fire, wrap garbage, swat flies, fill boxes, line drawers, make up kidnap note, for example) before the baseline task was performed and again before the stimulation task. Following Shamay-Tsoory and colleagues [21, 22], participants were asked to think of original uses of everyday objects. Participants were given up to one minute to complete each item of the task for a total of 5 minutes for each part of the experiment. The intraoperative testing was recorded and documented (following participants' consent) and blindly graded postoperatively. Measures of originality, based on a previously validated infrequency measure, and fluency, average number of valid creative uses given for each task item, were derived. For the fluency score, the number of alternative uses that did not recapitulate the original use (i.e., pencil—to write, to draw etc.) were counted. For the originality score, we utilized a previously constructed and validated answer database of 92 normal healthy adults. This allowed for a valid response frequency estimation. Each use was scored based on the infrequency in this answer database (score of 0 if >5% of participants provided it, 1 is 2–5%, 2 if <2%). For example—alternate uses for a shoe (common use to wear on feet) included "door stopper" 14.1%, score of 0 in originality, "make loud noises—bang with" 2.1% score of 1, and "put on the head for balance training" 1.08% score of 2. According to this, average originality and creative fluency scores were calculated for each patient and for each experimental condition. This scoring method is well validated and was chosen as it is objective, thus minimizing observer bias [23, 24]. In addition, this advantage is especially important as this study focused on the change in AUT score following DN stimulation, and did not attempt to predict future overall creative ability based on the baseline score. Both the participant and the examiner were blinded to the direct cortical stimulation location.

Intraoperative direct cortical stimulation

Direct cortical stimulation was performed using the Ojemann cortical stimulator (Radionics Inc., Burlington, MA, USA). For language mapping, current intensity was modified by increasing the amplitude in 2 mA increments, from a baseline of 2 mA up to 8 mA. If seizures were observed, ice water was irrigated over the brain, and the study was aborted. During cortical mapping for language functions, observed behavioral effects and performance dysfunctions (e.g., speech arrest, anomia, etc.) were tagged and documented with regard to both anatomical and radiological location, as well as to the applied current intensity. If no effect was produced following language mapping, the stimulator was set to 4 mA, and stimulation was performed for 10 sec followed by 10 sec without stimulation until the completion of the task item.

RESULTS

A total of 13 out of 22 eligible patients who underwent awake resection of a primary brain tumor were included in this study (Table 1). Nine patients were excluded from this study, six due to intraoperative clinical or electrocorticographic seizures and 3 due to speech impairments induced during language mapping of the exposed cortex.

All patients were high-functioning and cognitively intact (see table S1 for detailed cognitive assessment available for 12/13 patients) with no language disturbances at baseline following a preoperative neuropsychological evaluation. In addition, all patients were neurologically intact as documented in their hospital admission neurological examination. All patients had undergone assessment of a baseline creativity score of >0 in both AUT domains (originality and fluency of creative thinking), and

Table 1. Cohort demographics and AUT results.

#	Age	Gender	Tumor location	Pathology	Planned target stimulation site	Baseline originality	Baseline fluency	Stimulation originality	Stimulation fluency
1	24	F	Parietal	AA	DN	0.9	3	1.558333	2.085
2	26	M	Temporo-Insular	Astro	DN	0.48	3.6	0.81	2.88
3	60	M	Temporal	GBM	DN	0.66	1.6	0	1.5
4	69	M	Parietal	GBM	DN	0.83	3	0.418333	2.6
5	19	F	Frontal	Astro	DN	2	1	0.2	0.4
6	24	M	Temporal	Astro	DN	0.53	3.33	1	2.2
7	51	F	Insula	Oligo	DN	1.67	2	1.41	1.6
8	41	M	Frontal	Astro	DN	0.3	3.6	0.85	2.8
9	22	M	Temporo-Parietal	Astro	DN	0.5	2.667	0	1.5
10	34	M	Frontal	GBM	Control	0.4	1	0.625	2.125
11	32	M	Frontal	Oligo	Control	0.5	3.67	0.0975	1.1675
12	29	M	Temporal	Oligo	Control	0.43	3.75	0.35	4
13	20	M	Temporal	Astro	Control	0.55	2.33	1	0.5

F female, *M* male, *AA* Anaplastic Astrocytoma WHO grade III, *Astro* Astrocytoma WHO grade II, *GBM* Glioblastoma Multiforme WHO grade IV, *Oligo* Oligodendroglioma WHO grade II, *DN* default network.

thus were eligible to participate in this experiment. All patients willfully agreed to participate and signed an informed consent prior to surgery. Nine participants underwent stimulation of a DN node as derived from resting-state functional connectivity MRI posterior cingulate cortex (PCC) region of interest (ROI) seed-based whole-brain correlation map. Cases where the node was in proximity to the tumor, and inside the cortical region that was meant to be exposed according to a clinically-planned craniotomy participated in the experiment as subjects, and underwent stimulation of the node. Four participants who did not have an accessible DN node underwent stimulation of a non-DN cortical area and served as the control group. Importantly, prior to participation in this experiment, the exposed cortex was mapped intraoperatively for language functions as is the case with all awake craniotomies performed at our center. If any language disturbances were produced, or if a seizure (clinical or subclinical, as recorded by electrocorticography) was induced by the stimulation, the participant was excluded from this study (Fig. 1).

To validate AUT in the operating room setting, we correlated the baseline scores obtained during surgery (prior to stimulation) with each individual participant's correlation strength (z(r)) between PCC and the left medial prefrontal cortex (mPFC), a surrogate marker of DN integrity (*14*, Fig. 2A). Fluency was positively correlated with PCC-mPFC correlation ($r_{(11)} = 0.68$, $P = 0.007$, Fig. 2B), while originality was not ($r_{(11)} = -0.44$, $P = 0.11$, Fig. 2C).

We first examined the effects of stimulation, and stimulation site (DN node compared to loci outside of the DN) on creative thinking. A two-way ANOVA was conducted to examine the effects of DCS and Stimulation Site on the AUT Fluency Score. This analysis revealed an effect of DCS on Fluency, which was reduced following stimulation ($F(1,11) = 7.47$, $P = 0.019$), but none with Stimulation Site ($F(1,11) = 0$, $P = 0.99$). Surprisingly, no interaction was found between DCS and Stimulation Site ($F(1,11) = 0$, $P = 0.72$), suggesting a mismatch between preoperative planning and operative localization of the DN. A two-way ANOVA comparing the effects of DCS and Stimulation Site on Originality Score revealed no effects (Stimulation: $F(1,11) = 0$, $P = 0.99$; Stimulation Site: $F(1,11) = 0.79$, $P = 0.39$; Stimulation × Stimulation Site: $F(1,11) = 1.79$, $P = 0.21$). These findings indicate that in our cohort, DCS of the left hemisphere specifically affected creative fluency but not originality, and this effect is independent of planned stimulation site.

Next, in order to further validate these findings and to minimize the potential effects of anatomical or surgically-associated disruptions on stimulation accuracy, such as intraoperative brain shift, we conducted a post-hoc analysis in which a seed was placed at the final point of stimulation (as was recorded intraoperatively using the navigation system). These ROIs (Fig. 3A) were used to calculate individualized whole-brain correlation maps, which were then compared to the canonical DN as defined by Yeo et al. [3]. First, we validated that the group DN map as derived from a PCC seed region overlaps with the canonical DN map (Fig. 3B). Next, we hypothesized that in participants where the final stimulation location derived seed-based maps are more similar to canonical DN, DCS will yield a more prominent effect on creative thinking (see Fig. 3C for individual participant's examples of a stimulation ROI derived maps that show high (i) vs. low (ii) similarity to the canonical DN). Next, for each participant, we computed the overlap between the individualized stimulation-based map and the canonical DN map using the Sørensen-Dice similarity coefficient score (Fig. 3D). Then, we divided the cohort into two subgroups based on the Similarity score, and examined the effects of DCS in these two subgroups (Fig. 3E). A two-way ANOVA was conducted to examine the effect of Stimulation, and Similarity on the AUT Fluency Score. The analysis revealed that while Similarity did not have a direct effect on Fluency ($F_{(1,11)} = 0.02$, $P = 0.91$), it demonstrated a robust interaction with Stimulation ($F_{(1,11)} = 12.62$, $P = 0.005$), with a stronger effect of stimulation in individuals in the high Similarity subgroup. Importantly, examining this effect in individuals based on the anatomical location of the stimulation site (Fig. 3F), we found that the change in Fluency could be produced by stimulation of parietal, frontal and temporal regions, suggesting that it is network-dependent. In addition, no effect or interaction were found for Originality (Similarity: $F_{(1,11)} = 0.1$, $P = 0.76$; Stimulation × Similarity: $F_{(1,11)} = 0.56$, $P = 0.47$). Collectively, these findings suggest that stimulation in areas that are more connected to the DN (as defined by the canonical DN) produces a strong effect on fluency, but not originality.

DISCUSSION

Using a novel experimental approach—disruption of individually identified regions within the default network using direct cortical stimulation in awake patients performing a divergent thinking task—we demonstrate a causal link between the default network and creative fluency. Stimulation of left (dominant) hemisphere DN nodes had no effect on originality scores, and stimulation outside the DN did not cause any disruption to creative thinking. We further validated this finding by demonstrating that direct cortical stimulation-induced impairment of creative fluency was specific to and more prominent in participants who were stimulated in DN-associated regions.

Creative cognition, a hallmark of higher cognitive function, is essential for human evolution and progress. The ability to perform creative thinking in both abstract forms, such as art, and in complex problem solving, such as science, necessitates multi-modal information integration together with introspective thought processes. This manifests in the ability to link seemingly unrelated elements [25, 26]. These qualities may be explained by the unexpected synchronization of spatially segregated nodes of different computational modalities. Recent creative cognition models have postulated that creativity emerges through a synchronization between three cortical networks—the DN, salience and the executive control network [11, 27]. Each of these networks is considered to accommodate a different aspect of creative thinking. The DN, a network that mediates spontaneous cognition, or the "stream of consciousness", is thought to contribute to the flexible retrieval of memories and generation of ideas [7]. The salience network is thought to filter useful and novel candidate ideas and forward them to the executive control network that constrains this stream towards a specific goal [28]. Various studies have correlated DN integrity or strength of internal correlation between different hubs with creativity psychometrics. Takeuchi et al. showed that reduced task-induced deactivation of the PCC is associated with increased creativity, postulating that reallocation of cognitive resources to DN domains instead of working memory may be responsible for enhanced creative thinking [29]. Abraham et al. used a conceptual expansion fMRI task, a specific form of divergent thinking, and found that both salience and DN nodes are overactivated in participants with higher creativity, suggesting that individual differences are dependent on network integration [30]. Benedek et al. [31], using a block design fMRI AUT, examined the functional organization of divergent thinking. In their study of creative idea generation, a deactivation of ventral attention components, and of the PCC within the DN, was observed, while increased activity was noted in other DN nodes, such as the prefrontal cortex, and the hippocampal formation. These results suggest that different aspects of creativity are modulated by specialized DN subcomponents. In another work, examining chains of free associations, Marron et al. demonstrated substantial involvement of the DN in free association production compared to other forms of language production [32]. Interestingly, in their study the majority of reported activations were in the left hemisphere, providing a

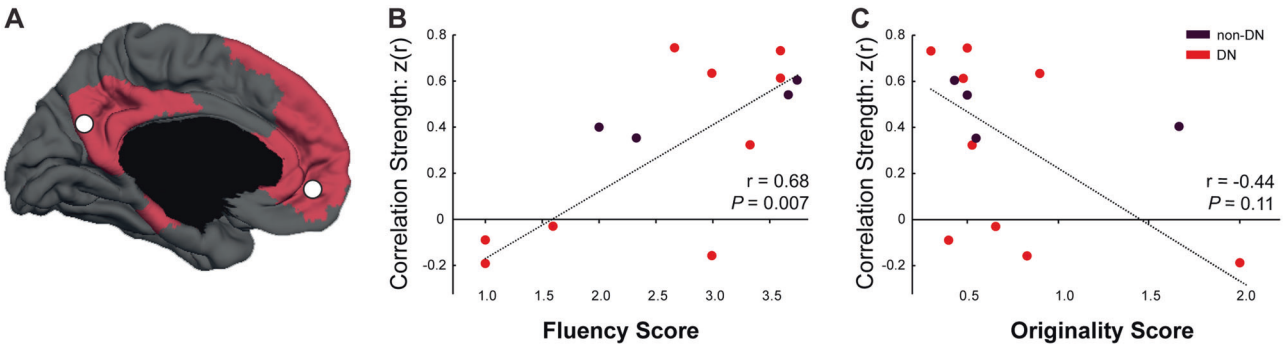


Fig. 2 Intraoperative performance in the Alternate Uses Task (AUT) relative to Default Network (DN) resting-state functional connectivity MRI strength. A For each participant, the correlations between two key nodes of the DN, PCC and mPFC were computed. The extent of DN coverage, derived from Yeo et al. [3], is depicted on the medial surface (red) along with the PCC and mPFC seed region-of-interest locations (denoted by white filled circles). **B** AUT fluency score at baseline (no stimulation condition) is correlated with PCC–mPFC Fisher's z-transformed Pearson's *r* correlation strength (*z*(*r*)). **C** AUT originality score is not correlated with PCC–mPFC correlation strength.

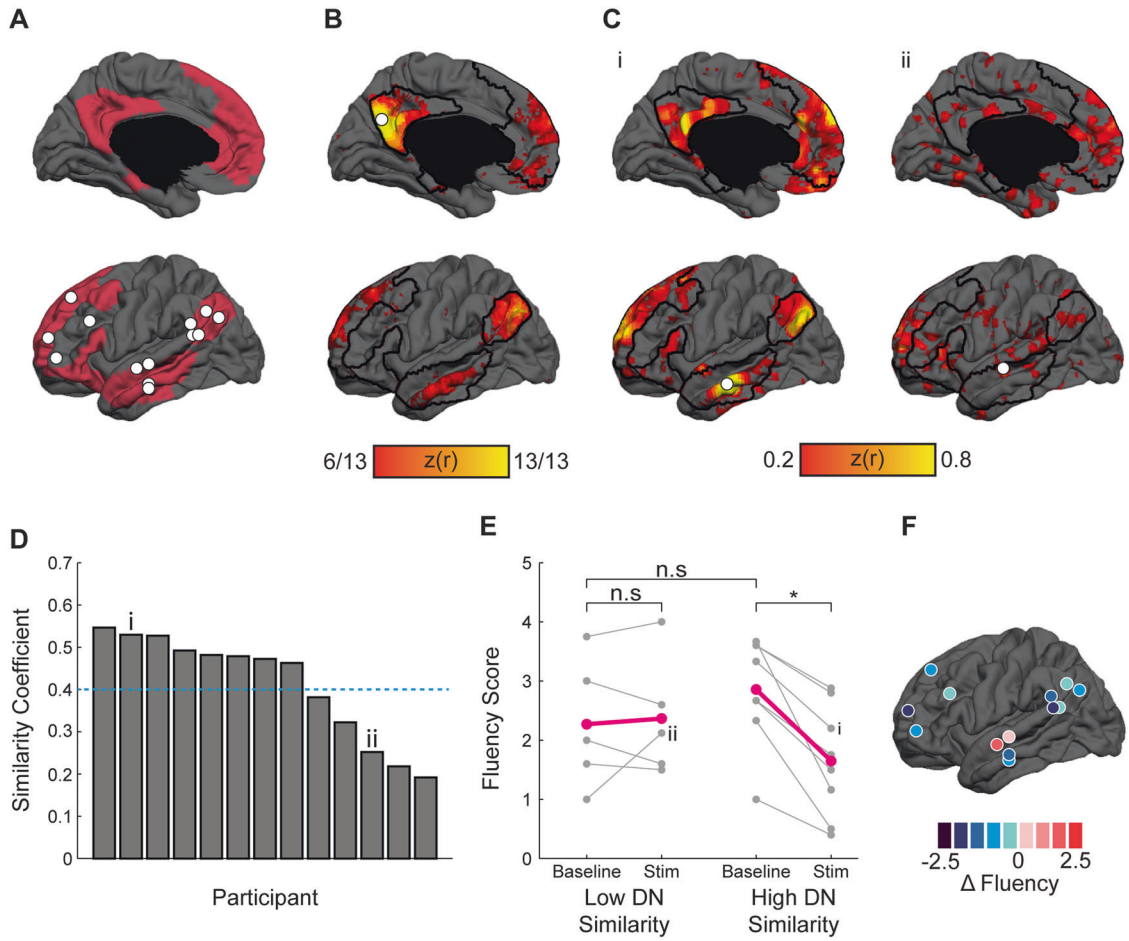


Fig. 3 Direct cortical stimulation of the DN impairs creative fluency. A DN spatial extent adopted from Yeo et al. [3] is depicted. Direct cortical stimulation sites in the DN ($n = 9$) and in control regions ($n = 4$) are depicted (filled white circles). **B** The entire cohort ($n = 13$) DN map was extracted using a PCC seed region (denoted by a filled white circle). The map shows the convergence of individual PCC maps thresholded at *z*(*r*) of 0.2. **C** Representative individual participants' maps derived from a seed region of interest within the DN placed at the stimulation site (filled white circle). A representative participant that demonstrates a map resembling the canonical DN (**i**), and a participant that does not (**ii**). **D** Individual participants overlap with the canonical DN (similarity coefficient) between the map obtained from the stimulation site ROI and the canonical DN. Participants i and ii are marked. **E** Participants with a similarity coefficient of 0.4 and above demonstrated significant impairment in creative fluency following direct cortical stimulation (*$P < 0.05$, n.s. non-significant; participants i and ii are marked, purple lines represent group average). **F** The relationship between direct cortical stimulation anatomical location and stimulation-induced changes in fluency are denoted as the difference in score change from baseline to the stimulation condition (filled dark-blue to dark-red circles).

possible explanation for our finding of decreased creative fluency following DN DCS. The emerging concept of sub-networks within the DN (for review see Buckner and DiNicola [33]) that are associated with various aspects of creativity is further supported by Shamay-Tsoory et al. [21]. This study, on which we relied methodologically, examined patients with localized lesions utilizing the same version of the AUT used here. Their finding of a correlation between right hemispheric lesions in the mPFC and decreased originality led them to associate this region with less linear cognitive processes. Interestingly, this finding is in line with our results demonstrating reduced creative fluency with DCS of left hemisphere DN and no change in originality, further suggesting a role for right hemisphere DN in original thinking. It should be noted that verbal and creative fluency are well correlated, and may partially overlap in function [34, 35]. However, in all participating subjects, the cortex exposed during surgery was thoroughly mapped for language functions (stimulated systematically while evaluating for naming, comprehension, and free speech). In all of the included subjects there was no disruption of speech-related functions, suggesting that the observed effect is creative fluency-specific, and not related to language disruption in general, further emphasizing the role of the DN in creative cognition. In addition, none of the patients were stimulated in the left inferior frontal gyrus, a main node in verbal fluency, which may disrupt both of these overlapping functions. In addition, recent evidence supports the existence of at least two DN subnetworks that are functionally specialized, and anatomically distinct in individual patients [36]. Two main subcomponents of the DN were recently characterized as mediating either theory of mind, or episodic projection [37]. This intra-DN parcellation may, in part, explain the variability in our participants' performance, as well as the fact that only some components of creativity were decreased by DCS of an individual's DN. An alternative explanation for the selective disruption of creative fluency seen following DN DCS is that the main contribution of the DN to divergent thinking is through mediation and retrieval of memories (for review see Hass and Beaty, 2018 [38]). This possible interference with episodic retrieval, is in line with recent evidence that boosting of episodic memory led to better performance in AUT [39]. As it is extremely rare for patients harboring gliomas in the non-dominant hemisphere to undergo awake resection, a complementary experiment adjudicating these two possible explanations (i.e., demonstrating reduced originality following right-sided DN DCS) was not performed in our particular setting, but is a subject for future research.

Several attempts were made over recent years to map, using electrical stimulation, what is termed "non-eloquent" cortex [40]. One such attempt by Foster et al. targeted the postero-medial cortex, a key node of the DN, via depth electrodes implanted in epilepsy patients [41]. In their study, and in concordance with our experience, when mapping areas of associative networks, no effect was evident. This lack of cognitive effects may be explained by the lack of an appropriate task or by the lack of patient-specific, network-based targets for stimulation. In our study design, the stimulated areas that yielded significant behavioral effects upon stimulation were those that produced a DN-like map at the individual level. Also, patients who demonstrated cognitive impairments or lack of creative ability in the preoperative neuropsychological assessment were excluded, as they were not expected to be able to fully participate in such a task.

We are aware of the limitations in our study design, as the number of subjects is limited and stimulation areas were limited to the left hemisphere. In addition, the number of task items and the time allowed for each item may influence the performance in AUT [42]. However the time frame in this special setting is extremely constrained [43], and we believe that this effect was minimized by the study design in which each participant served as their own control with no comparison of creative ability to an external cohort.

These limitations are a derivative of the unique experimental setting in which the human cortex is stimulated while an awake patient performs a task in the operating room. Recently, Natu et al. demonstrated that stimulation of the PCC in epilepsy patients implanted with depth electrodes impairs episodic memory recall, while no other subjective experience was elicited. In addition, following extended periods of stimulation of the PCC, a modulation of the PCC-hippocampal network, a key component of the DN, was evident [44]. These works, taken together with the findings presented here, set the stage for future interventional studies targeting associative networks-related functions that may greatly advance our understanding of the complex processes that lie at the core of human essence. These findings suggest a future approach for the mapping and preservation of creativity and higher cognitive functions in patients undergoing brain surgery.

REFERENCES

1. Jung RE, Segall JM, Jeremy Bockholt H, Flores RA, Smith SM, Chavez RS, et al. Neuroanatomy of creativity. Hum Brain Mapp. 2009;31:398–409.
2. Yuste R. From the neuron doctrine to neural networks. Nat Rev Neurosci. 2015;16:487–97.
3. Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol. 2011;106:1125–65.
4. Buckner RL, Krienen FM. The evolution of distributed association networks in the human brain. Trends Cogn Sci. 2013; 2013. https://doi.org/10.1016/j.tics.2013.09.017.
5. Rosenberg MD, Finn ES, Scheinost D, Papademetris X, Shen X, Constable RT, et al. A neuromarker of sustained attention from whole-brain functional connectivity. Nat Neurosci. 2016;19:165–71.
6. Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, et al. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat Neurosci. 2015;18:1664–71.
7. Andrews-Hanna JR, Reidler JS, Huang C, Buckner RL. Evidence for the default network's role in spontaneous cognition. J Neurophysiol. 2010;104:322–35.
8. Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: anatomy, function, and relevance to disease. Ann N Y Acad Sci. 2008;1124:1–38.
9. McKiernan KA, D'Angelo BR, Kaufman JN, Binder JR. Interrupting the "stream of consciousness": An fMRI investigation. NeuroImage. 2006;29:1185–91.
10. Wei D, Yang J, Li W, Wang K, Zhang Q, Qiu J. Increased resting functional connectivity of the medial prefrontal cortex in creativity by means of cognitive stimulation. Cortex. 2014;51:92–102.
11. Beaty RE, Kenett YN, Christensen AP, Rosenberg MD, Benedek M, Chen Q, et al. Robust prediction of individual creative ability from brain functional connectivity. Proc Natl Acad Sci USA. 2018;115:1087–92.
12. Thakral PP, Madore KP, Kalinowski SE, Schacter DL. Modulation of hippocampal brain networks produces changes in episodic simulation and divergent thinking. Proc. Natl Acad Sci USA. 2020;117:12729–40.
13. Thakral PP, Madore KP, Schacter DL. A role for the left angular gyrus in episodic simulation and memory. J Neurosci. 2017;37:8142–9.
14. Berger MS, Ojemann GA. Intraoperative brain mapping techniques in neuro-oncology. Stereotact Funct Neurosurg. 1992;58:153–61.
15. Guilford, JP, Christensen, PR, Merrifield, PR, & Wilson, RC. Alternate uses: Manual of instructions and interpretation. Orange, CA: Sheridan Psychological Services; 1978.
16. Doniger GM, Dwolatzky T, Zucker DM, Chertkow H, Crystal H, Schweiger A, et al. Computerized cognitive testing battery identifies mild cognitive impairment and mild dementia even in the presence of depressive symptoms. Am J Alzheimers Dis Other Demen. 2006;21:28–36.
17. Nossek E, Matot I, Shahar T, Barzilai O, Rapoport Y, Gonen T, et al. Intraoperative seizures during awake craniotomy: incidence and consequences: analysis of 477 patients. Neurosurgery. 2013;73:135–40.
18. Nossek E, Matot I, Shahar T, Barzilai O, Rapoport Y, Gonen T, et al. Failed awake craniotomy: a retrospective analysis in 424 patients undergoing craniotomy for brain tumor. J Neurosurg. 2013;118:243–9.
19. Shofty B, Bergmann E, Zur G, Asleh J, Bosak N, Kavushansky A, et al. Autism-associated Nf1 deficiency disrupts corticocortical and corticostriatal functional connectivity in human and mouse. Neurobiol Dis. 2019;130:104479.

20. Bergmann E, Zur G, Bershadsky G, Kahn I. The organization of mouse and human cortico-hippocampal networks estimated by intrinsic functional connectivity. Cereb Cortex. 2016;26:4497–512.
21. Shamay-Tsoory SG, Adler N, Aharon-Peretz J, Perry D, Mayseless N. The origins of originality: The neural bases of creative thinking and originality. Neuropsychologia. 2011;49:178–85.
22. Mayseless N, Eran A, Shamay-Tsoory SG. Generating original ideas: The neural underpinning of originality. Neuroimage. 2015;116:232–9.
23. Wallach MA, Kogan N. Modes of thinking in young children. New York: Holt, Rinehart and Winston; 1965.
24. van de Kamp M-T, Admiraal W, van Drie J, Rijlaarsdam G. Enhancing divergent thinking in visual arts education: Effects of explicit instruction of meta-cognition. Br J Educ Psychol. 2015;85:47–58.
25. James W. The principles of psychology. New York: Henry Holt and Company; 1890.
26. Spearman C, Charles D. Creative mind. J Nerv Ment Dis. 1931;74:783.
27. Beaty RE, Benedek M, Silvia PJ, Schacter DL. Creative cognition and brain network dynamics. Trends Cogn Sci (Regul Ed). 2016;20:87–95.
28. Chrysikou EG, Weber MJ, Thompson-Schill SL. A matched filter hypothesis for cognitive control. Neuropsychologia. 2014;62:341–55.
29. Takeuchi H, Taki Y, Hashizume H, Sassa Y, Nagase T, Nouchi R, et al. The association between resting functional connectivity and creativity. Cereb Cortex. 2012;22:2921–9.
30. Abraham A, Rutter B, Bantin T, Hermann C. Creative conceptual expansion: a combined fMRI replication and extension study to examine individual differences in creativity. Neuropsychologia. 2018;118:29–39.
31. Benedek M, Jauk E, Fink A, Koschutnig K, Reishofer G, Ebner F, et al. To create or to recall? Neural mechanisms underlying the generation of creative new ideas. NeuroImage. 2014;88:125–33.
32. Marron TR, Lerner Y, Berant E, Kinreich S, Shapira-Lichter I, Hendler T, et al. Chain free association, creativity, and the default mode network. Neuropsychologia. 2018;118:40–58.
33. Buckner RL, DiNicola LM. The brain's default network: updated anatomy, physiology and evolving insights. Nat Rev Neurosci. 2019;20:593–608.
34. Broday-Dvir R, Malach R. Resting-state fluctuations underlie free and creative verbal behaviors in the human brain. Cereb Cortex. 2021;31:213–32.
35. Silvia PJ, Beaty RE, Nusbaum EC. Verbal fluency and creativity: General and specific contributions of broad retrieval ability (Gr) factors to divergent thinking. Intelligence. 2013;41:328–40.
36. Gordon EM, Laumann TO, Marek S, Raut RV, Gratton C, Newbold DJ, et al. Default-mode network streams for coupling to language and control systems. Proc Natl Acad Sci USA. 2020. July 2020. https://doi.org/10.1073/pnas.2005238117.
37. DiNicola LM, Braga RM, Buckner RL. Parallel distributed networks dissociate episodic and social functions within the individual. J Neurophysiol. 2020;123:1144–79.
38. Hass RW, Beaty RE. Use or consequences: probing the cognitive difference between two measures of divergent thinking. Front Psychol. 2018;9:2327.
39. Madore KP, Jing HG, Schacter DL. Divergent creative thinking in young and older adults: extending the effects of an episodic specificity induction. Mem Cogn. 2016;44:974–88.
40. Fox KCR, Shi L, Baek S, Raccah O, Foster BL, Saha S, et al. Intrinsic network architecture predicts the effects elicited by intracranial electrical stimulation of the human brain. Nat Hum Behav. 2020. July 2020. https://doi.org/10.1038/s41562-020-0910-1.
41. Foster BL, Parvizi J. Direct cortical stimulation of human posteromedial cortex. Neurology. 2017;88:685–91.
42. Said-Metwaly S, Fernández-Castilla B, Kyndt E, Van den Noortgate W. Testing conditions and creative performance: Meta-analyses of the impact of time limits and instructions. Psychol Aesthet Creativity Arts. 2020;14:15–38.
43. Gonen T, Sela G, Yanakee R, Ram Z, Grossman R. Surgery-independent language function decline in patients undergoing awake craniotomy. World Neurosurg. 2017;99:674–9.
44. Natu VS, Lin J-J, Burks A, Arora A, Rugg MD, Lega B. Stimulation of the posterior cingulate cortex impairs episodic memory encoding. J Neurosci. 2019;39:7173–82.

AUTHOR CONTRIBUTIONS

Authors BS, TG, NM, SST, ZR conceived the experiment; BS, TG, EB, NM, SST, IK, ZR designed the analysis; BS, TG, AK, RG, IJ, ZR collected the data; BS, TG, EB, IJ performed the analysis; BS, TG, IK, ZR wrote the paper; EB, NM, AK, SST, RG, IJ critically reviewed the manuscript.

COMPETING INTERESTS

The author declares no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-021-01403-8.

Correspondence and requests for materials should be addressed to Ben Shofty.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

Central and Peripheral Circadian Clocks in Mammals

Jennifer A. Mohawk,[1] Carla B. Green,[1] and Joseph S. Takahashi[1,2]

[1]Department of Neuroscience, [2]Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9111;
email: Jennifer.Mohawk@UTSouthwestern.edu, Carla.Green@UTSouthwestern.edu, Joseph.Takahashi@UTSouthwestern.edu

Annu. Rev. Neurosci. 2012. 35:445–62

First published online as a Review in Advance on April 5, 2012

The *Annual Review of Neuroscience* is online at neuro.annualreviews.org

This article's doi:
10.1146/annurev-neuro-060909-153128

Copyright © 2012 by Annual Reviews.
All rights reserved

0147-006X/12/0721-0445$20.00

Keywords

clock genes, suprachiasmatic nucleus, oscillator coupling, metabolism, temperature

Abstract

The circadian system of mammals is composed of a hierarchy of oscillators that function at the cellular, tissue, and systems levels. A common molecular mechanism underlies the cell-autonomous circadian oscillator throughout the body, yet this clock system is adapted to different functional contexts. In the central suprachiasmatic nucleus (SCN) of the hypothalamus, a coupled population of neuronal circadian oscillators acts as a master pacemaker for the organism to drive rhythms in activity and rest, feeding, body temperature, and hormones. Coupling within the SCN network confers robustness to the SCN pacemaker, which in turn provides stability to the overall temporal architecture of the organism. Throughout the majority of the cells in the body, cell-autonomous circadian clocks are intimately enmeshed within metabolic pathways. Thus, an emerging view for the adaptive significance of circadian clocks is their fundamental role in orchestrating metabolism.

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

Contents

INTRODUCTION 446
MOLECULAR MECHANISM OF THE CIRCADIAN CLOCK IN MAMMALS 447
CENTRAL CIRCADIAN OSCILLATORS 448
Suprachiasmatic Nucleus 448
PERIPHERAL CLOCKS 451
ORGANIZATION OF THE CIRCADIAN SYSTEM 451
Neural Control of Peripheral Oscillators: The Autonomic Nervous System 452
Hormonal Control of Peripheral Oscillators 453
Temperature 453
Behavioral and Homeostatic Regulation: Local Cues Feed Back into the Clock 454
OTHER OSCILLATORS: FOOD AND DRUGS 455
The Food-Entrainable Oscillator 455
The Methamphetamine-Sensitive Circadian Oscillator 456
SUMMARY 457

INTRODUCTION

Circadian rhythm: an endogenously generated oscillation with a period of ~24 h; can entrain to external cues and is temperature compensated

Suprachiasmatic nucleus (SCN): the site of the master circadian pacemaker in mammals, comprising ~20,000 individual neurons that couple to form a robust oscillatory network

Living systems possess an exquisitely accurate internal biological clock that times daily events ranging from sleep and wakefulness in humans to photosynthesis in plants (Takahashi et al. 2008). These circadian rhythms represent an evolutionarily conserved adaptation to the environment that can be traced back to the earliest life forms. In animals, circadian behavior can be analyzed as an integrated system—beginning with genes and leading ultimately to behavioral outputs. In the past 15 years, the molecular mechanism of circadian clocks has been uncovered by the use of phenotype-driven (forward) genetic analysis in a number of model systems (Lowrey & Takahashi 2011). Circadian oscillations are generated by a set of genes forming a transcriptional autoregulatory feedback loop. In mammals, these include *Clock*, *Bmal1*, *Per1*, *Per2*, *Cry1*, and *Cry2*. Researchers have identified another dozen candidate genes that play additional roles in the circadian gene network such as the feedback loop involving *Rev-erbα*.

Early work on mammalian rhythms used rhythmic behavior as a readout of the clock, and the hypothalamic suprachiasmatic nucleus (SCN) was identified as the dominant circadian pacemaker driving behavioral rhythms (Welsh et al. 2010). However, the discovery of "clock genes" led to the realization that the capacity for circadian gene expression is widespread throughout the body (Dibner et al. 2010). Using circadian gene reporter methods, one can demonstrate that most peripheral organs and tissues can express circadian oscillations in isolation yet still receive, and may require, input from the SCN in vivo. The cell-autonomous clock is ubiquitous, and almost every cell in the body contains a circadian clock (Balsalobre et al. 1998, Nagoshi et al. 2004, Welsh et al. 2004, Yoo et al. 2004). It is now evident that the circadian oscillators within individual cells respond differently to entraining signals, control different physiological outputs, and interact with each other and with the system as a whole. These discoveries have raised a number of questions concerning the synchronization and coherence of rhythms at the cellular level as well as the architecture of circadian clocks at the systems level (Hogenesch & Ueda 2011). Here we discuss recent work that addresses these organizational issues and examines a number of levels of complexity within the circadian system. We review mechanisms by which circadian clocks govern biological processes as well as mechanisms by which these processes feed back into the circadian system. Perhaps the most important example of this is the intimate and reciprocal interaction between the circadian clock system and fundamental metabolic pathways (Bass & Takahashi 2010, Green et al. 2008, Rutter et al. 2002). In addition, there exist additional oscillatory processes in the circadian time domain that are observable in the presence of scheduled meals or methamphetamine

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

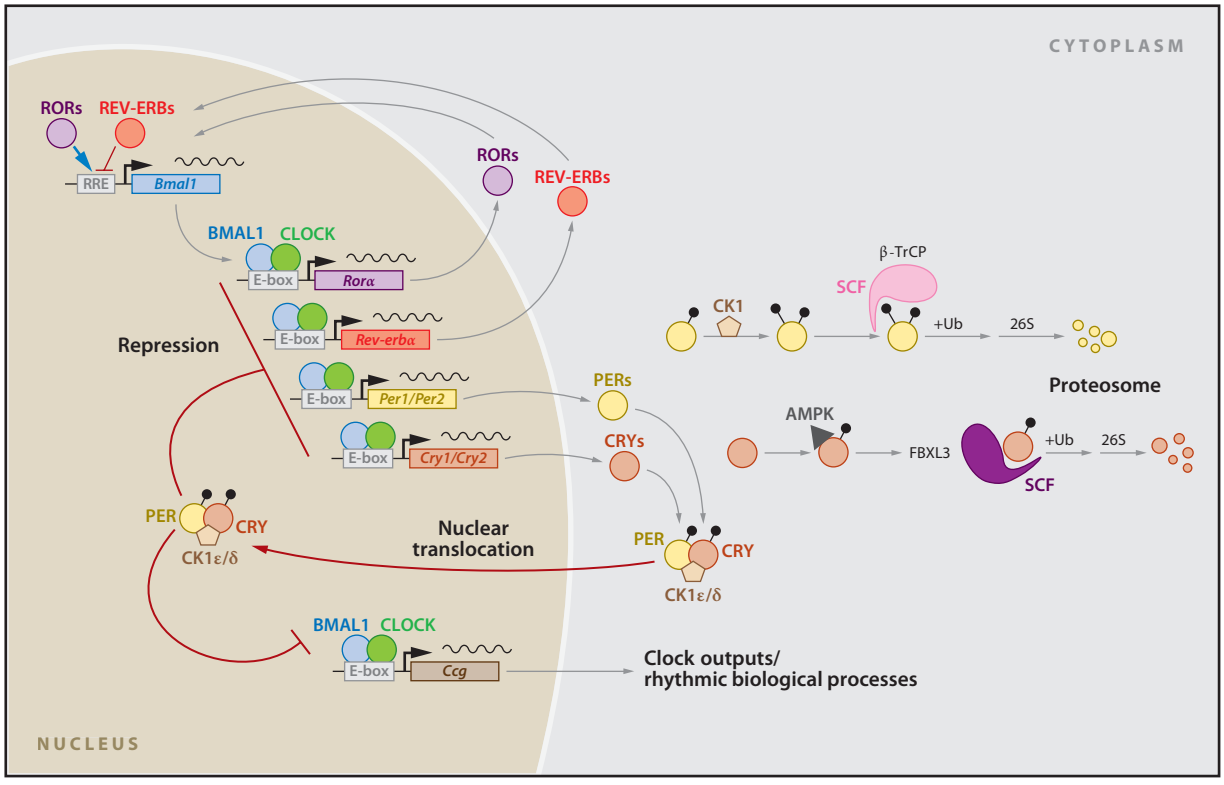


Figure 1

The molecular mechanism of the circadian clock in mammals. Constituting the core circadian clock is an autoregulatory transcriptional feedback loop involving the activators CLOCK and BMAL1 and their target genes *Per1*, *Per2*, *Cry1*, and *Cry2*, whose gene products form a negative-feedback repressor complex. In addition to this core transcriptional feedback loop, other feedback loops are also driven by CLOCK:BMAL1. One feedback loop involving *Rev-erbα* and *Rorα* that represses *Bmal1* transcription leads to an antiphase oscillation in *Bmal1* gene expression. CLOCK:BMAL1 also regulates many downstream target genes known as clock-controlled genes (Ccg). At a post-transcriptional level, the stability of the PER and CRY proteins is regulated by SCF (Skp1-Cullin-F-box protein) E3 ubiquitin ligase complexes involving β-TrCP and FBXL3, respectively. The kinases, casein kinase 1ε/δ (CK1ε/δ) and AMP kinase (AMPK), phosphorylate the PER and CRY proteins, respectively, to promote polyubiquitination by their respective E3 ubiquitin ligase complexes, which in turn tag the PER and CRY proteins for degradation by the 26S proteasome complex.

treatment. These oscillators can generate behavioral rhythms in vivo in the absence of the SCN (Honma & Honma 2009).

MOLECULAR MECHANISM OF THE CIRCADIAN CLOCK IN MAMMALS

In mammals, the mechanism of the circadian clock is cell autonomous and arises from an autoregulatory negative-feedback transcriptional network (Lowrey & Takahashi 2004, Takahashi et al. 2008) (**Figure 1**). At the core of this clock network are the transcriptional activators, CLOCK (and its paralog, NPAS2) and BMAL1, which positively regulate the expression of the *Period* (*Per1*, *Per2*) and *Cryptochrome* (*Cry1*, *Cry2*) genes at the beginning of the cycle. *Per* and *Cry* gene products accumulate, dimerize, and form a complex that translocates into the nucleus to interact with CLOCK and BMAL1, repressing their own transcription. This feedback cycle takes ~24 h, and the turnover of the PER and CRY proteins is tightly regulated by E3 ubiquitin ligase complexes. There are additional feedback loops interlocked

with the core CLOCK-BMAL1/PER-CRY loop. Prominent among these is a loop involving *Rev-erbα* (*Nr1d1*) and *Rora*, which are also direct targets of CLOCK-BMAL1. The feedback effects of this loop impinge on the transcription of *Bmal1* (and to a lesser extent on *Clock*) to cause an antiphase oscillation of BMAL1. Other feedback loops involve the PAR-bZip family members, DBP, HLF, and TEF; the bZip protein, E4BP4 (*Nfil3*); and the bHLH proteins, DEC1 and DEC2 (*Bhlhb2*, *Bhlhb3*), all of which are transcriptional targets of CLOCK-BMAL1 (Gachon 2007, Lowrey & Takahashi 2004, Takahashi et al. 2008).

The discovery of a ubiquitous, cell-autonomous clock in mammals has led to a re-evaluation of central and peripheral oscillators: Are they fundamentally similar in mechanism, how do they function in different cellular contexts, and what role does coupling in the central SCN clock play in its functional properties?

CENTRAL CIRCADIAN OSCILLATORS

The hypothalamic SCN acts as a master pacemaker for the generation of circadian behavioral rhythms in mammals (for a review, see Welsh et al. 2010). Classic work not reviewed here has shown that the SCN is both necessary and sufficient for the generation of circadian activity rhythms in rodents. The SCN receives direct photic input from the retina from a recently discovered photoreceptor cell type termed the intrinsically photoreceptive retinal ganglion cell (ipRGC) (reviewed in Do & Yau 2010). These ipRGCs express a novel photopigment, melanopsin, that renders them intrinsically photosensitive to short-wavelength irradiation. Interestingly, ipRGCs are depolarizing photoreceptors that employ a phototransduction mechanism that is analogous to that seen in invertebrate photoreceptors. The photoresponse in ipRGCs has slow kinetics and a relatively high threshold to light, making them ideally suited to function as circadian photoreceptors, which must integrate light information over relatively long durations and must be insensitive to transient light signals that are not associated with the solar light cycle. Although ipRGCs appear to be optimal circadian photoreceptors, they do not act alone; rod and cone photoreceptors also have photic inputs to the SCN. Interestingly, these nonvisual inputs from rods and cones to the SCN are mediated by the ipRGCs (Chen et al. 2011, Guler et al. 2008). An emerging theme is that melanopsin-positive ipRGCs are involved in a surprisingly broad array of nonvisual photic responses in mammals. The complexity of the ipRGCs and their contribution to circadian rhythms and other behaviors are beyond the scope of this discussion, but recent reviews have covered this topic in depth (Do & Yau 2010, Schmidt et al. 2011).

Suprachiasmatic Nucleus

The SCN is composed of ~20,000 neurons, each of which is thought to contain a cell-autonomous circadian oscillator. The SCN functions as a network in which the population of SCN cells are coupled together and oscillate in a coherent manner (Herzog 2007). The dynamics of the spatial and temporal coordination of rhythms in the SCN have been studied recently, enabled by the advent of single-cell circadian reporter technology, which has revealed unexpected complexity in the temporal architecture of the nucleus (Evans et al. 2011, Foley et al. 2011, Yamaguchi et al. 2003). At the single-cell level, SCN neurons exhibit a wide range in cell-autonomous circadian periods that vary from 22 h to 30 h (Ko et al. 2010, Liu et al. 1997, Welsh et al. 1995). Intercellular coupling among SCN neurons acts to mutually couple the entire population to a much narrower range that corresponds to the circadian period of the locomotor activity rhythm, which is extremely precise (a standard deviation in period that is ~0.2 h or 12 min in mice) (Herzog et al. 2004). The heterogeneity in intrinsic period of the SCN cells confers at least two important functions: phase lability and phase plasticity. The phases of the rhythms of individual SCN neurons are highly stereotyped

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

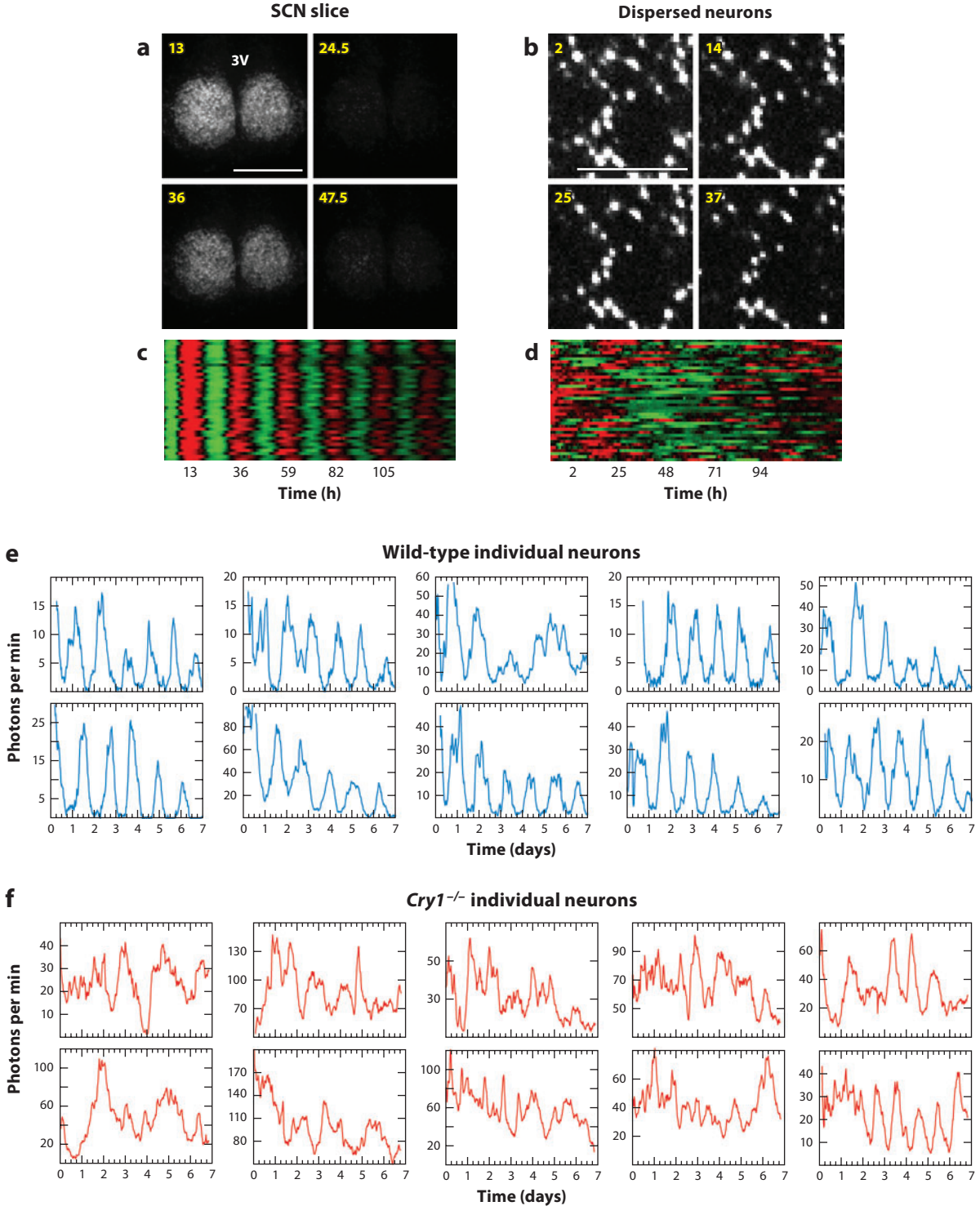
anatomically and appear as a wave that spreads across the nucleus over time (Evans et al. 2011, Foley et al. 2011, Yamaguchi et al. 2003). Intrinsically shorter-period cells have earlier phases and intrinsically longer-period cells have later phases within the SCN, reflecting phase lability (Yamaguchi et al. 2003). Under different photoperiods (e.g., long- versus short-photoperiod light cycles), the waveform of the SCN population rhythm is modulated such that in short photoperiods the SCN waveform is narrow and has a high amplitude, whereas in long photoperiods the SCN waveform is broad and has a low amplitude, reflecting phase plasticity (Inagaki et al. 2007, VanderLeest et al. 2007). In addition to the heterogeneity of SCN oscillator period and phase, it has been proposed that the cell-autonomous SCN oscillators are not intrinsically uniformly robust (Webb et al. 2009). Rather, the intercellular coupling of SCN neurons appears critical to the robustness of the SCN network oscillatory system.

With the discovery of peripheral oscillators (Balsalobre et al. 1998, Yamazaki et al. 2000, Yoo et al. 2004) and the apparent ubiquity of clock mechanisms (Yagita et al. 2001), a critical question arises concerning the similarity and differences in the SCN pacemaker as compared with peripheral oscillators. To address this question, Liu et al. (2007a) examined whether canonical clock mutations previously assessed in vivo affected the SCN and peripheral oscillators in a similar manner. Using *Per2::luciferase* reporter mice, they found that the effects of the *Period* and *Cryptochrome* loss-of-function mutations were the same in SCN explants as those seen previously at the behavioral level. By contrast, in peripheral tissues, single loss-of-function mutations that are subtle at the behavioral level, such as *Per1* or *Cry1* knockouts, produced very strong loss-of-rhythm phenotypes. Interestingly, the effects of these mutations are cell autonomous in both fibroblasts and in isolated SCN neurons, supporting the idea that the cell-autonomous clock is similar in these two cell types. However, when the SCN population is coupled, the effects of these mutations are non-cell autonomous. This occurs as a consequence of the intercellular coupling in the SCN network, which is capable of rescuing a cell-autonomous defect in the individual cells (**Figure 2**). This transformation of the oscillatory capability of SCN neurons from damped to self-sustained is an important illustration of the robustness of the SCN network. Indeed, Ko et al. (2010) have found that the SCN network is capable of generating oscillations in the circadian domain in the complete absence of cell-autonomous oscillatory potential. In *Bmal1*-knockout mice, which are arrhythmic at the behavioral level, SCN explants unexpectedly express stochastic oscillations in the circadian range that are highly variable. When the individual cells are no longer rhythmic, the coupling pathways within the SCN network can propagate stochastic rhythms that are a reflection of both feed-forward coupling mechanisms and intracellular noise. Thus, in a manner analogous to central pattern generators in neural circuits, rhythmicity can arise as an emergent property of the network in the absence of the component pacemaker or oscillator cells.

In addition to the generation of sustained oscillations by the SCN network, the SCN is also robust to perturbations from environmental inputs. In wild-type mice, the phase-resetting curve to light pulses is characteristic of Type 1 or weak resetting (low amplitude) (Vitaterna et al. 2006). This is a reflection of the robustness of the SCN pacemaker because inputs such as light can perturb the phase of the oscillation only to a limited extent. In contrast, genetic mutations that lower the amplitude of the molecular oscillation in the SCN lead to increases in the sensitivity to light-induced phase shifts (Type 0 resetting) without changing the strength of the light signals impinging on the SCN (Vitaterna et al. 2006). Similar effects are seen with temperature cycles. Peripheral oscillators are exquisitely sensitive to the phase-shifting effects of temperature and can be entrained strongly by low-amplitude temperature cycles that are equivalent to the circadian fluctuation in core body temperature

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.



(~2.5°C oscillation in mice) (Brown et al. 2002, Buhr et al. 2010). Interestingly, the SCN is resistant to entrainment by low-amplitude temperature cycles, and this resistance depends on the intercellular coupling of SCN neurons (Abraham et al. 2010, Buhr et al. 2010). As was the case for genetic mutations on the generation of rhythms, the effects of temperature perturbations on the SCN are cell autonomous when intercellular coupling is eliminated. Thus, both SCN and peripheral oscillators are sensitive to temperature cycles at the cell-autonomous level; however, coupling within the SCN network confers robustness and makes the SCN network resistant to temperature perturbations. As discussed below, the temperature resistance of the SCN makes functional sense because the SCN drives the body-temperature rhythm, and this temperature signal, which can serve as an entraining signal for peripheral clocks, may otherwise feed back and interfere with the SCN (Buhr et al. 2010).

PERIPHERAL CLOCKS

Rhythms of clock gene and/or protein expression have been observed in cells and tissues throughout the body in mammals, and these rhythms persist in culture, demonstrating that non-SCN cells also contain endogenous circadian oscillators (Balsalobre et al. 1998, Yamazaki et al. 2000, Yoo et al. 2004). Although the core clock machinery is conserved in these different cellular clocks, there are significant differences in the relative contributions of the individual clock components, as well as in the manner in which these peripheral clocks are reset and in the output pathways that are under their control. These endogenous cellular clocks drive extensive rhythms of gene transcription, with 3–10% of all mRNAs in a given tissue showing circadian rhythms in steady-state levels (Akhtar et al. 2002, Duffield et al. 2002, Hughes et al. 2009, Miller et al. 2007, Panda et al. 2002, Storch et al. 2002). However, the genes that are under circadian control are largely nonoverlapping in each tissue, reflecting the need for temporal control of the cellular physiology relevant to each unique cell type. As a result, the circadian clock exerts broad-ranging control over many biological processes, including many aspects of metabolism such as xenobiotic detoxification (Gachon et al. 2006), glucose homeostasis (Lamia et al. 2008, Marcheva et al. 2010, So et al. 2009, Turek et al. 2005), and lipogenesis (Gachon et al. 2011, Le Martelot et al. 2009).

Entrainment: synchronization and phase control of a rhythm to a regularly occurring environmental cycle (usually ~24 h)

ORGANIZATION OF THE CIRCADIAN SYSTEM

For biological clocks to be effective, they must accurately keep time and adjust to environmental signals. In an organized circadian system, this requires SCN control of peripheral oscillators, and loss of the SCN results in peripheral circadian clocks that become desynchronized (Yoo et al. 2004). However, tissue-specific gene expression patterns are likely to be regulated by both "local" as well as central mechanisms. This concept was elegantly demonstrated through genetic disruption of the

Figure 2

Network and autonomous properties of suprachiasmatic nucleus (SCN) neurons. Network properties of the SCN can compensate for genetic defects affecting rhythmicity at the cell-autonomous level. (*a*) Bioluminescence images of a *Cry1*$^{-/-}$ SCN in organotypic slice culture. Note the stable, synchronized oscillations. Numbers indicate hours after start of imaging; 3V indicates the third ventricle. (*b*) Bioluminescence images of dissociated individual *Cry1*$^{-/-}$ SCN neurons showing cell-autonomous, largely arrhythmic patterns of high bioluminescence intensity. (*c*,*d*) Heat-map representations of bioluminescence intensity of individual *Cry1*$^{-/-}$ neurons in an (*a*) SCN slice and (*b*) dispersed culture. Values above and below the mean are shown in red and green, respectively, for 40 SCN neurons in each condition. (*e*,*f*) Ten single SCN neuron rhythms from (*e*) wild-type and (*f*) *Cry1*$^{-/-}$ mice. Imaging began immediately following a media change at day 0. Dissociated *Cry1*$^{-/-}$ SCN neurons are largely arrhythmic, whereas dissociated wild-type cells are rhythmic. By contrast, in organotypic slice cultures, both wild-type and *Cry1*$^{-/-}$ SCN cells are robustly rhythmic and tightly synchronized. Figure and legend adapted and reprinted from Liu et al. (2007a), with permission from Elsevier.

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

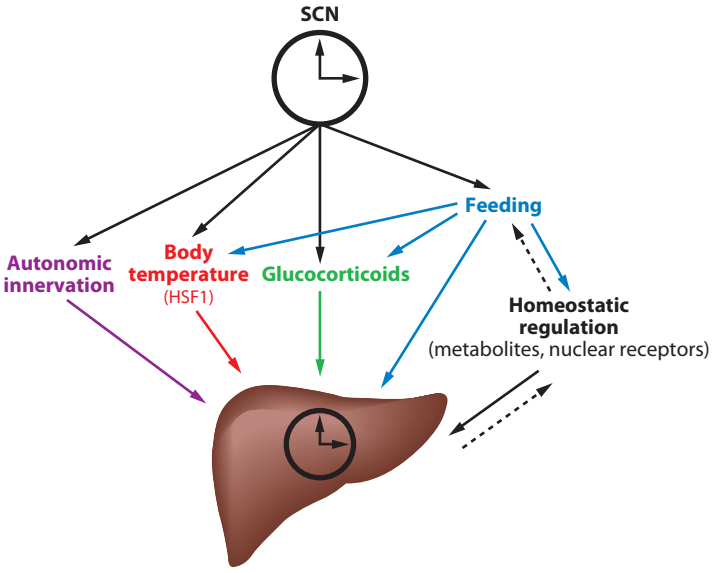


Figure 3

Pathways of peripheral clock entrainment. The master circadian pacemaker within the suprachiasmatic nucleus (SCN) relays temporal information to peripheral oscillators through autonomic innervation, body temperature, humoral signals (such as glucocorticoids), and feeding-related cues. Independent of the SCN, local signaling pathways can also affect peripheral oscillators.

circadian clock mechanism specifically in the hepatocytes of mice, while leaving the circadian clock intact in the SCN and other cell types throughout the body (Kornmann et al. 2007). Microarray analysis of mRNAs in the livers from these mice demonstrated that the disruption of the circadian molecular feedback loop specifically within the liver results in arrhythmicity of most hepatic transcripts. Thus, most circadian oscillations of hepatic function rely on an intact liver clock. However, a subset of transcripts, including the core clock component *Per2*, continued to cycle robustly even in the absence of a functional liver clock. In livers maintained in explant culture, rhythms in *Per2* transcription were observed in livers with intact clocks but were absent in livers with inactivated clocks. Thus, rhythmic gene expression can be driven by both local intracellular clocks and by extracellular systemic cues.

What are these systemic cues? The photically entrained SCN is thought to convey signals to light-insensitive peripheral clocks to synchronize these systems, and SCN transplant studies (Ralph et al. 1990, Silver et al. 1996) and parabiosis experiments in mice (Guo et al. 2005) have demonstrated that both humoral and nonhumoral pathways are important for SCN coordination of circadian output rhythms. In addition, complex feedback loops link the circadian clock with rhythmic metabolic networks, integrating these systems in a light-independent manner. Circadian control of metabolism occurs at the central (SCN) as well as local levels and involves clocks within a number of peripheral tissues including the liver, pancreas, skeletal muscle, intestine, and adipose tissue (for a review, see Bass & Takahashi 2010, Green et al. 2008). This intimate relationship between clocks and metabolism is an example of how the circadian "system" is integrated with, and influenced by, the physiology that is under its control. Therefore, organization of the circadian system requires a combination of (*a*) autonomic innervation of peripheral tissues, (*b*) endocrine signaling, (*c*) temperature, and (*d*) local signals (**Figure 3**).

Neural Control of Peripheral Oscillators: The Autonomic Nervous System

The SCN controls peripheral oscillators through both sympathetic and parasympathetic pathways (Kalsbeek et al. 2010, Ueyama et al. 1999). SCN projections through the paraventricular nucleus–superior cervical ganglia (PVN-SCG) pathway provide the dominant entraining signal for the submandibular salivary glands (Ueyama et al. 1999, Vujovic et al. 2008). Sympathetic innervation from the SCN to the PVN to the liver results in daily rhythms of plasma glucose, presumably by directly influencing the rhythm of hepatic gluconeogenesis (Cailotto et al. 2005, Kalsbeek et al. 2004).

Autonomic pathways from the SCN relay photic information to oscillators in the adrenal gland and liver (Buijs et al. 1999, Cailotto et al. 2009, Ishida et al. 2005). Sympathetic innervation also modulates the sensitivity of the adrenal to adrenocorticotropic hormone (ACTH) and

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

directly influences glucocorticoid release (Buijs et al. 1999, Kalsbeek et al. 2010, Kaneko et al. 1981). Oscillators in both the adrenal cortex and medulla respond to neural inputs emanating from the SCN (Buijs et al. 1999, Mahoney et al. 2010). The adrenal clock is of particular interest given the strong case for glucocorticoids as a humoral entraining signal for peripheral clocks.

Hormonal Control of Peripheral Oscillators

Although a number of hormones may have roles in mammalian circadian organization, glucocorticoids have received the most attention. Rhythmic glucocorticoids result both from the sympathetic inputs discussed above and from an underlying rhythm of corticotropin releasing hormone (CRH) and ACTH function (Kaneko et al. 1980, Kaneko et al. 1981). The adrenal clock also provides temporal control of sensitivity to ACTH-induced glucocorticoid release (Oster et al. 2006).

The demonstration that dexamethasone (a glucocorticoid analog) could shift the phase of peripheral tissues in vivo (Balsalobre et al. 2000) provided the first definitive evidence that glucocorticoids were entraining signals for peripheral oscillators. Dexamethasone was initially shown to shift the phase of clock gene expression in the liver, kidney, and heart as well as cultured fibroblasts. What provides glucocorticoid input into the clock at the local level? Glucocorticoid-response elements exist in the regulatory regions of the core clock genes, *Bmal1*, *Cry1*, *Per1* (Reddy et al. 2007, Yamamoto et al. 2005), and *Per2* (So et al. 2009). These glucocorticoid-response elements may lead to the transcriptional activation of a number of clock genes and clock-controlled genes by glucocorticoids.

Glucocorticoids can synchronize circadian expression of much of the oscillatory component of the liver transcriptome in SCN-lesioned mice (Reddy et al. 2007). This is accomplished, in part, through activation of the nuclear receptor (and hepatic transcription factor) HNF4α. HNF4α is responsive to glucocorticoids (Reddy et al. 2007), contains E-boxes, which may allow for transcriptional control by CLOCK:BMAL1 (Reddy et al. 2007), and can interact with PER2 (Schmutz et al. 2010). Other metabolically relevant nuclear receptors, including peroxisome proliferator-activated receptor (PPAR)α, also respond to glucocorticoids (Lemberger et al. 1994). Nuclear receptor activation by clock components and glucocorticoids provides another point of circadian input to metabolic pathways as described below.

PPAR: peroxisome proliferator-activated receptor

HSF1: heat shock factor 1

Temperature

In most organisms, temperature is a powerful entraining agent for circadian rhythms. However, in mammals, external temperature cycles are very weak entraining agents (Refinetti 2010); this has been attributed to the fact that homeotherms regulate their body temperature and can defend their body temperature against environmental fluctuations. It has long been known that body temperature is circadian and the rhythm is driven by the SCN. Peripheral oscillators, including fibroblasts, liver, kidney, and lung, are exquisitely sensitive to temperature changes (Abraham et al. 2010, Brown et al. 2002, Buhr et al. 2010, Kornmann et al. 2007). These oscillators can be strongly reset by low-amplitude temperature pulses that mimic the range of circadian variation, and temperature profiles that match circadian body temperature rhythms strongly entrain peripheral clocks (Brown et al. 2002, Buhr et al. 2010). However, as described above, the SCN is resistant to temperature cycles in the circadian range. Because of this system design, the SCN is ideally situated to utilize circadian temperature cycles as a universal entraining signal for peripheral oscillators. The influence of temperature on peripheral oscillators likely occurs through the transcription factor heat shock factor 1 (HSF1). HSF1 transcriptional activity oscillates with a circadian rhythm in the liver and can be driven by temperature cycles (Reinke et al. 2008). HSF1 inhibitors block temperature-induced resetting in extra-SCN

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

NAD: nicotinamide adenine dinucleotide

oscillators (Buhr et al. 2010). Because HSF1 is influenced by a wide range of signaling pathways in the cell (Akerfelt et al. 2010), temperature and HSF1 may form a final common pathway for the integration of resetting signals in peripheral clocks.

Behavioral and Homeostatic Regulation: Local Cues Feed Back into the Clock

In addition to controlling hormone secretion and body temperature directly, the SCN coordinates rhythms in behavioral processes, such as locomotor activity and feeding, which can influence endocrine function and body temperature. These behaviors, feeding in particular, can regulate peripheral clocks at the local level, modulating local signaling pathways and metabolic processes. Homeostatic signaling pathways also affect peripheral clocks and their function, allowing for extra SCN control of circadian processes. Studies in which mealtime has been experimentally manipulated to occur antiphase to the normal SCN-driven feeding rhythm have attempted to elucidate local mechanisms for controlling clocks in peripheral tissues.

The liver clock, unlike the SCN, is particularly sensitive to resetting by feeding. Hepatic rhythms of clock gene and protein expression rapidly shift their phase to follow the timing of a scheduled meal (Damiola et al. 2000, Stokkan et al. 2001). Similarly, livers of *Cry1/Cry2*-null mice display rhythms in many transcripts (including a number of transcripts involved in metabolic processes) when fed in regular 24-h intervals (Vollmers et al. 2009). Feeding appears to result in cues that bypass the core circadian feedback loop to drive these rhythms. These cues may include feeding-induced changes in temperature and HSF1 activity (Kornmann et al. 2007) or activation of other metabolically sensitive pathways.

A number of local mediators of both core clock components and clock-controlled rhythmic transcripts have been identified, and can respond to SCN-driven inputs as well as local signals related to homeostasis and metabolic state. These mediators include members of the nuclear receptor family of transcription factors, many of which exhibit circadian rhythms of transcription within the liver and other metabolically relevant tissues (Yang et al. 2006). These rhythmic nuclear receptors regulate transcription of downstream metabolic pathways. Among the rhythmic nuclear receptors are PPARs and members of the REV-ERB and ROR families. As described above, RORα and REV-ERBα participate directly in the clock mechanism by regulating *Bmal1* transcription (Preitner et al. 2002, Sato et al. 2004), but they are also important for many aspects of metabolic regulation. Similar to REV-ERBs and RORs, many of the other rhythmic nuclear receptors are regulators of clock function, providing a mechanism by which signals of metabolic status can influence rhythmicity. Glucocorticoid receptors, as discussed above, induce transcription of *Per* and potentially a number of other clock and clock-controlled genes (Reddy et al. 2007, So et al. 2009, Yamamoto et al. 2005). PPARα, which responds to lipid status and glucocorticoids, may also regulate *Bmal1* transcription (Canaple et al. 2006).

PPARγ coactivator-1α (PGC-1α, a transcriptional coactivator) provides a link between the clock and changes in metabolic status. PGC-1α is critical for adaptive responses to nutritional and metabolic state, particularly following fasting (reviewed in Lin et al. 2005). PGC-1α is rhythmic and activates expression of *Bmal1* and *Rev-erbα* through coactivation of RORs (Liu et al. 2007b). PGC-1α-null mice display disruptions in a number of circadian outputs including locomotor activity, oxygen consumption rate, and expression of both clock and metabolic genes (Liu et al. 2007b). PGC-1α also interacts with SIRTUIN 1 (SIRT1), a nicotinamide adenine dinucleotide (NAD)-dependent histone deacetylase (Rodgers et al. 2005).

Another mechanism by which metabolic signals can feed into the clock is through the adenosine monophosphate-activated protein kinase (AMPK) (Bass & Takahashi 2010). This kinase is a central mediator of metabolic

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

signals. AMPK activity is robustly rhythmic in mouse liver and is regulated by nutrient status, as reflected in the ratio of AMP to ATP. Active AMPK directly regulates the central clock mechanism by phosphorylating and destabilizing the clock component CRY1 (Lamia et al. 2009).

Cellular redox state can also serve as a mechanism by which the metabolic status of the cell can impact the circadian system. NAD levels exhibit circadian oscillations in the liver, likely owing to transcriptional regulation of nicotinamide phosphoribosyltransferase (*Nampt*, encoding the rate-limiting enzyme in the NAD^+ salvage pathway) by CLOCK:BMAL1 (Nakahata et al. 2009, Ramsey et al. 2009). NAD levels also vary with cellular redox state as a consequence of metabolic changes, which, in turn, can directly impact clock function. The ratio of NAD^+ to NADH influences binding of NPAS2:BMAL1 and CLOCK:BMAL1 to DNA in vitro, suggesting one way in which NAD could interact with clock components (Rutter et al. 2001).

NAD^+-dependent SIRT1 also displays daily oscillations and feeds back onto the circadian clock. SIRT1 forms a complex with CLOCK:BMAL1, leading to the deacetylation of PER2 (Asher et al. 2008) and BMAL1 (Nakahata et al. 2008). SIRT1 also suppresses CLOCK:BMAL1-mediated transcription, resulting in decreased expression of *Per2* (Nakahata et al. 2008, Ramsey et al. 2009) and the clock-controlled gene *Dbp* (Nakahata et al. 2008).

Another molecule with NAD^+-sensitive activity, poly(ADP-ribose) polymerase 1 (PARP-1), was recently shown to interact with the circadian clock (Asher et al. 2010). PARP-1 activity is rhythmic in the liver, and this rhythm persists even in the absence of a functional hepatic circadian clock. The rhythm of PARP-1 activity can be entrained by scheduled meals, however, suggesting that the circadian activity of PARP-1 is driven by feeding-related cues. PARP-1 interacts with CLOCK:BMAL1 in a rhythmic fashion and inhibits DNA binding by the CLOCK:BMAL1 complex. PARP-1 also polyADP-ribosylates CLOCK and appears to temporally regulate the interaction of CLOCK:BMAL1 with PER2 and CRYs. The circadian regulation of PARP-1 by feeding, and subsequent consequences for the circadian clock, are likely not entirely mediated through NAD^+. Regardless of the mechanism, PARP-1 provides another way for metabolic signals to influence timekeeping by the core molecular clock.

OTHER OSCILLATORS: FOOD AND DRUGS

The circadian system consists of a web of interconnected oscillators and feedback loops. The core molecular clock within cells keeps time and responds to cues from the SCN (through neural, hormonal, and activity-driven pathways) as well as signals from the local cellular environment. These cell- and tissue-level clocks result in rhythms of physiologically relevant outputs, including glucose production, fat storage, and hormone production. These outputs, in turn, become circadian time-keeping cues relayed to other clocks throughout the body, likely ultimately feeding back to the central nervous system and the SCN. Under normal circumstances, the SCN maintains temporal organization of body temperature, activity, feeding, and neural output rhythms. This keeps local and systemic circadian signals aligned. In the absence of the SCN, however, the system becomes disorganized. Activity is nonrhythmic and peripheral tissues and cells drift out of phase with one another. There are two striking exceptions to this phenomenon. Scheduled, restricted feeding and chronic administration of methamphetamine, a psychostimulant drug of abuse, are both capable of organizing the circadian outputs in the absence of the SCN.

The Food-Entrainable Oscillator

It is not surprising that food and food-related cues are salient for many biological processes, including circadian rhythms. The ability of animals to anticipate food availability is well

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

established and persists even when food is provided at a time that is out of phase with the animal's normal feeding time (Richter 1922, Stephan et al. 1979). When food is temporally restricted to the daytime (the normal rest period), nocturnal rodents will anticipate the arrival of the meal with an increase in activity; if the timing of that meal is shifted, rats will display transients, gradually shifting their food-anticipatory activity bout each day until it again precedes the start of food availability.

In animals with lesions of the SCN, temporal food restriction will induce circadian rhythmicity of locomotor behavior (Stephan et al. 1979) and an accompanying temperature rhythm (Krieger et al. 1977). Food-anticipatory activity persists on days of total food deprivation, demonstrating that these cycles are not merely an hourglass phenomenon but are driven by an underlying oscillator (Stephan 2002). This food-entrainable oscillator (FEO) can take on pacemaking functions—organizing rhythms of activity, body temperature, and peripheral tissues in SCN-lesioned animals. Peripheral tissues from both SCN-intact and SCN-ablated mice are sensitive to temporally restricted feeding. In SCN-intact animals, phase desynchrony among peripheral oscillators can occur: Some tissues remain in phase with the (food-unaffected) SCN, and some follow the phase of food availability (Damiola et al. 2000, Pezuk et al. 2010). Cues related to the meal must be the dominant entraining signals in this latter group of tissues. In SCN-lesioned mice, food entrainment organizes rhythms throughout the periphery, and stable phase relationships are observed among tissues (Hara et al. 2001, Pezuk et al. 2010).

Interestingly, the FEO does not appear to require a functional molecular clock, as *Bmal1*$^{-/-}$ and *Per1/Per2*$^{-/-}$ mice can entrain to restricted feeding (Pendergast et al. 2009, Storch & Weitz 2009). The mechanism by which food drives oscillatory behavior throughout the organism is unknown. It is possible that the FEO exploits some of the same pathways used by the SCN, such as hormone- and temperature-dependent cues, to organize peripheral tissues. The locus (or loci) of the FEO is also unknown. A number of structures, including the olfactory bulbs (Davidson et al. 2001), the ventromedial hypothalamus (Mistlberger & Rechtschaffen 1984), the paraventricular thalamic nucleus (Landry et al. 2007), and a large portion of the digestive system (Davidson et al. 2003), have been ruled out.

The dorsomedial hypothalamus (DMH) has received considerable attention for its role in food entrainment. Data supporting a role for the DMH in the generation of food-anticipatory circadian activity are controversial (Gooley et al. 2006, Landry et al. 2006, Moriya et al. 2009), and the DMH may not be essential for the expression of the FEO (Landry et al. 2006, Moriya et al. 2009). The DMH does, however, interact with the SCN under conditions of food restriction and may influence the strength of the FEO output, particularly in SCN-intact animals (Acosta-Galvan et al. 2011).

The Methamphetamine-Sensitive Circadian Oscillator

Chronic or scheduled methamphetamine treatment affects circadian outputs in a manner similar to food restriction (Honma & Honma 2009). Methamphetamine, provided in the drinking water of rats and mice, is capable of driving circadian rhythms of locomotor behavior in the absence of the SCN (Honma et al. 1987, Tataroglu et al. 2006). In SCN-intact animals, this appears as a lengthening of the free-running period of locomotor activity, and in some cases, two activity components (relatively coordinated with each other) are observed. Much like the FEO, these rhythms persist when the stimulus (in this case, methamphetamine) is withdrawn (Tataroglu et al. 2006) as well as in the absence of a functional molecular clock (Honma et al. 2008, Mohawk et al. 2009). The methamphetamine-sensitive circadian oscillator (MASCO) is capable of functioning as a pacemaker driving rhythms in locomotor activity, body temperature, endocrine function, and the oscillators of peripheral

FEO: food-entrainable oscillator

DMH: dorsomedial hypothalamus

Free-running period: the period of an oscillation in the absence of entraining signals; reflects the intrinsic period of the oscillator uninfluenced by environmental timing cues

MASCO: methamphetamine-sensitive circadian oscillator

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

tissues (Honma et al. 1988, Pezuk et al. 2010). In the presence of the SCN, methamphetamine results in desynchrony among internal oscillators, as some follow the SCN and some follow the presumed phase of the MASCO (Pezuk et al. 2010). When the SCN is ablated, however, the MASCO organizes oscillators in tissues throughout the organism, resulting in a coordinated system (Pezuk et al. 2010).

The site of the MASCO is also unknown. It is possible that the FEO and MASCO share an anatomical and mechanistic basis, or, indeed, that they represent a single oscillator. Research focused on understanding how the MASCO and FEO relay circadian information to peripheral tissues will likely uncover novel (or underappreciated) mechanisms that control circadian rhythms. The role of these oscillators in the absence of food restriction and methamphetamine must also be determined. It is unlikely that these oscillators are dormant under normal conditions; instead, the FEO, MASCO, and SCN probably cooperate in a hierarchically organized, perhaps necessarily redundant, timing network.

SUMMARY

It is now clear that there is feedback at nearly every level of the circadian system. "Outputs" such as body temperature and feeding become inputs to other oscillators and are capable of influencing the core molecular clockwork, generating complex interconnectivity between the circadian system and the biological outputs it controls. Reciprocity between the circadian and metabolic systems makes it likely that perturbations in one system affect the other. This idea is supported both genetically—circadian mutants have metabolic phenotypes—and environmentally—nutrient intake can modulate circadian rhythms (Bass & Takahashi 2010). In recent years, considerable progress has been made in unraveling the connections between the circadian clock and metabolism. The role of circadian clocks in governing many other physiological systems has been established, but is far less well characterized.

We still know very little about how oscillators and timing cues are integrated at the local and organismal levels to coordinate the circadian architecture of the animal. Peripheral clocks must balance (sometimes conflicting) inputs arising from the SCN with those signaling local cellular and metabolic state. Moreover, recent work has revealed that the cell-autonomous oscillator, which normally lies at the foundation of the circadian clockwork, is not absolutely crucial for the expression of rhythms by other components of the system. Within the SCN, coupling among individual neurons gives rise to a heterogeneous, yet elegantly organized, robust oscillatory network, which can overcome impaired rhythmicity at the cellular level (Ko et al. 2010, Liu et al. 2007a). Food- and drug-sensitive oscillators (FEO, MASCO) can influence circadian rhythms and drive rhythmic outputs in the absence of the core molecular clock mechanism (Honma et al. 2008, Mohawk et al. 2009, Pendergast et al. 2009, Storch & Weitz 2009). The ability of rhythmic circadian outputs to persist in the absence of the SCN necessitates that any model of the circadian network include alternative mechanisms for controlling circadian rhythms at the cell, tissue, and organism levels.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holding that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Abraham U, Granada AE, Westermark PO, Heine M, Kramer A, Herzel H. 2010. Coupling governs entrainment range of circadian clocks. *Mol. Syst. Biol.* 6:438

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

Acosta-Galvan G, Yi CX, van der Vliet J, Jhamandas JH, Panula P, et al. 2011. Interaction between hypothalamic dorsomedial nucleus and the suprachiasmatic nucleus determines intensity of food anticipatory behavior. *Proc. Natl. Acad. Sci. USA* 108:5813–18

Akerfelt M, Morimoto RI, Sistonen L. 2010. Heat shock factors: integrators of cell stress, development and lifespan. *Nat. Rev. Mol. Cell Biol.* 11:545–55

Akhtar RA, Reddy AB, Maywood ES, Clayton JD, King VM, et al. 2002. Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven by the suprachiasmatic nucleus. *Curr. Biol.* 12:540–50

Asher G, Gatfield D, Stratmann M, Reinke H, Dibner C, et al. 2008. SIRT1 regulates circadian clock gene expression through PER2 deacetylation. *Cell* 134:317–28

Asher G, Reinke H, Altmeyer M, Gutierrez-Arcelus M, Hottiger MO, Schibler U. 2010. Poly(ADP-ribose) polymerase 1 participates in the phase entrainment of circadian clocks to feeding. *Cell* 142:943–53

Balsalobre A, Brown SA, Marcacci L, Tronche F, Kellendonk C, et al. 2000. Resetting of circadian time in peripheral tissues by glucocorticoid signaling. *Science* 289:2344–47

Balsalobre A, Damiola F, Schibler U. 1998. A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* 93:929–37

Bass J, Takahashi JS. 2010. Circadian integration of metabolism and energetics. *Science* 330:1349–54

Brown SA, Zumbrunn G, Fleury-Olela F, Preitner N, Schibler U. 2002. Rhythms of mammalian body temperature can sustain peripheral circadian clocks. *Curr. Biol.* 12:1574–83

Buhr ED, Yoo SH, Takahashi JS. 2010. Temperature as a universal resetting cue for mammalian circadian oscillators. *Science* 330:379–85

Buijs RM, Wortel J, Van Heerikhuize JJ, Feenstra MG, Ter Horst GJ, et al. 1999. Anatomical and functional demonstration of a multisynaptic suprachiasmatic nucleus adrenal (cortex) pathway. *Eur. J. Neurosci.* 11:1535–44

Cailotto C, La Fleur SE, Van Heijningen C, Wortel J, Kalsbeek A, et al. 2005. The suprachiasmatic nucleus controls the daily variation of plasma glucose via the autonomic output to the liver: Are the clock genes involved? *Eur. J. Neurosci.* 22:2531–40

Cailotto C, Lei J, van der Vliet J, van Heijningen C, van Eden CG, et al. 2009. Effects of nocturnal light on (clock) gene expression in peripheral organs: a role for the autonomic innervation of the liver. *PLoS One* 4:e5650

Canaple L, Rambaud J, Dkhissi-Benyahya O, Rayet B, Tan NS, et al. 2006. Reciprocal regulation of brain and muscle Arnt-like protein 1 and peroxisome proliferator-activated receptor alpha defines a novel positive feedback loop in the rodent liver circadian clock. *Mol. Endocrinol.* 20:1715–27

Chen SK, Badea TC, Hattar S. 2011. Photoentrainment and pupillary light reflex are mediated by distinct populations of ipRGCs. *Nature* 476:92–95

Damiola F, Le Minh N, Preitner N, Kornmann B, Fleury-Olela F, Schibler U. 2000. Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes Dev.* 14:2950–61

Davidson AJ, Aragona BJ, Werner RM, Schroeder E, Smith JC, Stephan FK. 2001. Food-anticipatory activity persists after olfactory bulb ablation in the rat. *Physiol. Behav.* 72:231–35

Davidson AJ, Poole AS, Yamazaki S, Menaker M. 2003. Is the food-entrainable circadian oscillator in the digestive system? *Genes Brain Behav.* 2:32–39

Dibner C, Schibler U, Albrecht U. 2010. The mammalian circadian timing system: organization and coordination of central and peripheral clocks. *Annu. Rev. Physiol.* 72:517–49

Do MT, Yau KW. 2010. Intrinsically photosensitive retinal ganglion cells. *Physiol. Rev.* 90:1547–81

Duffield GE, Best JD, Meurers BH, Bittner A, Loros JJ, Dunlap JC. 2002. Circadian programs of transcriptional activation, signaling, and protein turnover revealed by microarray analysis of mammalian cells. *Curr. Biol.* 12:551–57

Evans JA, Leise TL, Castanon-Cervantes O, Davidson AJ. 2011. Intrinsic regulation of spatiotemporal organization within the suprachiasmatic nucleus. *PLoS One* 6:e15869

Foley NC, Tong TY, Foley D, Lesauter J, Welsh DK, Silver R. 2011. Characterization of orderly spatiotemporal patterns of clock gene activation in mammalian suprachiasmatic nucleus. *Eur. J. Neurosci.* 33:1851–65

Gachon F. 2007. Physiological function of PARbZip circadian clock-controlled transcription factors. *Ann. Med.* 39:562–71

Gachon F, Leuenberger N, Claudel T, Gos P, Jouffe C, et al. 2011. Proline- and acidic amino acid-rich basic leucine zipper proteins modulate peroxisome proliferator-activated receptor alpha (PPARα) activity. *Proc. Natl. Acad. Sci. USA* 108:4794–99

Gachon F, Olela FF, Schaad O, Descombes P, Schibler U. 2006. The circadian PAR-domain basic leucine zipper transcription factors DBP, TEF, and HLF modulate basal and inducible xenobiotic detoxification. *Cell Metab.* 4:25–36

Gooley JJ, Schomer A, Saper CB. 2006. The dorsomedial hypothalamic nucleus is critical for the expression of food-entrainable circadian rhythms. *Nat. Neurosci.* 9:398–407

Green CB, Takahashi JS, Bass J. 2008. The meter of metabolism. *Cell* 134:728–42

Guler AD, Ecker JL, Lall GS, Haq S, Altimus CM, et al. 2008. Melanopsin cells are the principal conduits for rod-cone input to nonimage-forming vision. *Nature* 453:102–5

Guo H, Brewer JM, Champhekar A, Harris RB, Bittman EL. 2005. Differential control of peripheral circadian rhythms by suprachiasmatic-dependent neural signals. *Proc. Natl. Acad. Sci. USA* 102:3111–16

Hara R, Wan K, Wakamatsu H, Aida R, Moriya T, et al. 2001. Restricted feeding entrains liver clock without participation of the suprachiasmatic nucleus. *Genes Cells* 6:269–78

Herzog ED. 2007. Neurons and networks in daily rhythms. *Nat. Rev. Neurosci.* 8:790–802

Herzog ED, Aton SJ, Numano R, Sakaki Y, Tei H. 2004. Temporal precision in the mammalian circadian system: a reliable clock from less reliable neurons. *J. Biol. Rhythms* 19:35–46

Hogenesch JB, Ueda HR. 2011. Understanding systems-level properties: timely stories from the study of clocks. *Nat. Rev. Genet.* 12:407–16

Honma K, Honma S. 2009. The SCN-independent clocks, methamphetamine and food restriction. *Eur. J. Neurosci.* 30:1707–17

Honma K, Honma S, Hiroshige T. 1987. Activity rhythms in the circadian domain appear in suprachiasmatic nuclei lesioned rats given methamphetamine. *Physiol. Behav.* 40:767–74

Honma S, Honma K, Shirakawa T, Hiroshige T. 1988. Rhythms in behaviors, body temperature and plasma corticosterone in SCN lesioned rats given methamphetamine. *Physiol. Behav.* 44:247–55

Honma S, Yasuda T, Yasui A, van der Horst GT, Honma K. 2008. Circadian behavioral rhythms in Cry1/Cry2 double-deficient mice induced by methamphetamine. *J. Biol. Rhythms* 23:91–94

Hughes ME, DiTacchio L, Hayes KR, Vollmers C, Pulivarthy S, et al. 2009. Harmonics of circadian gene transcription in mammals. *PLoS Genet.* 5:e1000442

Inagaki N, Honma S, Ono D, Tanahashi Y, Honma K. 2007. Separate oscillating cell groups in mouse suprachiasmatic nucleus couple photoperiodically to the onset and end of daily activity. *Proc. Natl. Acad. Sci. USA* 104:7664–69

Ishida A, Mutoh T, Ueyama T, Bando H, Masubuchi S, et al. 2005. Light activates the adrenal gland: timing of gene expression and glucocorticoid release. *Cell Metab.* 2:297–307

Kalsbeek A, Bruinstroop E, Yi CX, Klieverik LP, La Fleur SE, Fliers E. 2010. Hypothalamic control of energy metabolism via the autonomic nervous system. *Ann. N.Y. Acad. Sci.* 1212:114–29

Kalsbeek A, La Fleur S, Van Heijningen C, Buijs RM. 2004. Suprachiasmatic GABAergic inputs to the paraventricular nucleus control plasma glucose concentrations in the rat via sympathetic innervation of the liver. *J. Neurosci.* 24:7604–13

Kaneko M, Hiroshige T, Shinsako J, Dallman MF. 1980. Diurnal changes in amplification of hormone rhythms in the adrenocortical system. *Am. J. Physiol.* 239:R309–16

Kaneko M, Kaneko K, Shinsako J, Dallman MF. 1981. Adrenal sensitivity to adrenocorticotropin varies diurnally. *Endocrinology* 109:70–75

Ko CH, Yamada YR, Welsh DK, Buhr ED, Liu AC, et al. 2010. Emergence of noise-induced oscillations in the central circadian pacemaker. *PLoS Biol.* 8:e1000513

Kornmann B, Schaad O, Bujard H, Takahashi JS, Schibler U. 2007. System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. *PLoS Biol.* 5:e34

Krieger DT, Hauser H, Krey LC. 1977. Suprachiasmatic nuclear lesions do not abolish food-shifted circadian adrenal and temperature rhythmicity. *Science* 197:398–99

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

Lamia KA, Sachdeva UM, DiTacchio L, Williams EC, Alvarez JG, et al. 2009. AMPK regulates the circadian clock by cryptochrome phosphorylation and degradation. *Science* 326:437–40

Lamia KA, Storch KF, Weitz CJ. 2008. Physiological significance of a peripheral tissue circadian clock. *Proc. Natl. Acad. Sci. USA* 105:15172–77

Landry GJ, Simon MM, Webb IC, Mistlberger RE. 2006. Persistence of a behavioral food-anticipatory circadian rhythm following dorsomedial hypothalamic ablation in rats. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 290:R1527–34

Landry GJ, Yamakawa GR, Mistlberger RE. 2007. Robust food anticipatory circadian rhythms in rats with complete ablation of the thalamic paraventricular nucleus. *Brain Res.* 1141:108–18

Le Martelot G, Claudel T, Gatfield D, Schaad O, Kornmann B, et al. 2009. REV-ERBalpha participates in circadian SREBP signaling and bile acid homeostasis. *PLoS Biol.* 7:e1000181

Lemberger T, Staels B, Saladin R, Desvergne B, Auwerx J, Wahli W. 1994. Regulation of the peroxisome proliferator-activated receptor alpha gene by glucocorticoids. *J. Biol. Chem.* 269:24527–30

Lin J, Handschin C, Spiegelman BM. 2005. Metabolic control through the PGC-1 family of transcription coactivators. *Cell Metab.* 1:361–70

Liu AC, Welsh DK, Ko CH, Tran HG, Zhang EE, et al. 2007a. Intercellular coupling confers robustness against mutations in the SCN circadian clock network. *Cell* 129:605–16

Liu C, Li S, Liu T, Borjigin J, Lin JD. 2007b. Transcriptional coactivator PGC-1α integrates the mammalian clock and energy metabolism. *Nature* 447:477–81

Liu C, Weaver DR, Strogatz SH, Reppert SM. 1997. Cellular construction of a circadian clock: period determination in the suprachiasmatic nuclei. *Cell* 91:855–60

Lowrey PL, Takahashi JS. 2004. Mammalian circadian biology: elucidating genome-wide levels of temporal organization. *Annu. Rev. Genomics Hum. Genet.* 5:407–41

Lowrey PL, Takahashi JS. 2011. Genetics of circadian rhythms in Mammalian model organisms. *Adv. Genet.* 74:175–230

Mahoney CE, Brewer D, Costello MK, Brewer JM, Bittman EL. 2010. Lateralization of the central circadian pacemaker output: a test of neural control of peripheral oscillator phase. *Am. J. Physiol.* 299:R751–61

Marcheva B, Ramsey KM, Buhr ED, Kobayashi Y, Su H, et al. 2010. Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature* 466:627–31

Miller BH, McDearmon EL, Panda S, Hayes KR, Zhang J, et al. 2007. Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. *Proc. Natl. Acad. Sci. USA* 104:3342–47

Mistlberger RE, Rechtschaffen A. 1984. Recovery of anticipatory activity to restricted feeding in rats with ventromedial hypothalamic lesions. *Physiol. Behav.* 33:227–35

Mohawk JA, Baer ML, Menaker M. 2009. The methamphetamine-sensitive circadian oscillator does not employ canonical clock genes. *Proc. Natl. Acad. Sci. USA* 106:3519–24

Moriya T, Aida R, Kudo T, Akiyama M, Doi M, et al. 2009. The dorsomedial hypothalamic nucleus is not necessary for food-anticipatory circadian rhythms of behavior, temperature or clock gene expression in mice. *Eur. J. Neurosci.* 29:1447–60

Nagoshi E, Saini C, Bauer C, Laroche T, Naef F, Schibler U. 2004. Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell* 119:693–705

Nakahata Y, Kaluzova M, Grimaldi B, Sahar S, Hirayama J, et al. 2008. The NAD^+-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control. *Cell* 134:329–40

Nakahata Y, Sahar S, Astarita G, Kaluzova M, Sassone-Corsi P. 2009. Circadian control of the NAD^+ salvage pathway by CLOCK-SIRT1. *Science* 324:654–57

Oster H, Damerow S, Kiessling S, Jakubcakova V, Abraham D, et al. 2006. The circadian rhythm of glucocorticoids is regulated by a gating mechanism residing in the adrenal cortical clock. *Cell Metab.* 4:163–73

Panda S, Antoch MP, Miller BH, Su AI, Schook AB, et al. 2002. Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* 109:307–20

Pendergast JS, Nakamura W, Friday RC, Hatanaka F, Takumi T, Yamazaki S. 2009. Robust food anticipatory activity in BMAL1-deficient mice. *PLoS One* 4:e4860

Pezuk P, Mohawk JA, Yoshikawa T, Sellix MT, Menaker M. 2010. Circadian organization is governed by extra-SCN pacemakers. *J. Biol. Rhythms* 25:432–41

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

Preitner N, Damiola F, Lopez-Molina L, Zakany J, Duboule D, et al. 2002. The orphan nuclear receptor REV-ERBα controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* 110:251–60

Ralph MR, Foster RG, Davis FC, Menaker M. 1990. Transplanted suprachiasmatic nucleus determines circadian period. *Science* 247:975–78

Ramsey KM, Yoshino J, Brace CS, Abrassart D, Kobayashi Y, et al. 2009. Circadian clock feedback cycle through NAMPT-mediated NAD^+ biosynthesis. *Science* 324:651–54

Reddy AB, Maywood ES, Karp NA, King VM, Inoue Y, et al. 2007. Glucocorticoid signaling synchronizes the liver circadian transcriptome. *Hepatology* 45:1478–88

Refinetti R. 2010. Entrainment of circadian rhythm by ambient temperature cycles in mice. *J. Biol. Rhythms* 25:247–56

Reinke H, Saini C, Fleury-Olela F, Dibner C, Benjamin IJ, Schibler U. 2008. Differential display of DNA-binding proteins reveals heat-shock factor 1 as a circadian transcription factor. *Genes Dev.* 22:331–45

Richter C. 1922. A behavioristic study of the activity of the rat. *Comp. Psychol. Monogr.* 1:1–55

Rodgers JT, Lerin C, Haas W, Gygi SP, Spiegelman BM, Puigserver P. 2005. Nutrient control of glucose homeostasis through a complex of PGC-1α and SIRT1. *Nature* 434:113–18

Rutter J, Reick M, McKnight SL. 2002. Metabolism and the control of circadian rhythms. *Annu. Rev. Biochem.* 71:307–31

Rutter J, Reick M, Wu LC, McKnight SL. 2001. Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors. *Science* 293:510–14

Sato TK, Panda S, Miraglia LJ, Reyes TM, Rudic RD, et al. 2004. A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. *Neuron* 43:527–37

Schmidt TM, Chen SK, Hattar S. 2011. Intrinsically photosensitive retinal ganglion cells: many subtypes, diverse functions. *Trends Neurosci.* 34:572–80

Schmutz I, Ripperger JA, Baeriswyl-Aebischer S, Albrecht U. 2010. The mammalian clock component PERIOD2 coordinates circadian output by interaction with nuclear receptors. *Genes Dev.* 24:345–57

Silver R, LeSauter J, Tresco PA, Lehman MN. 1996. A diffusible coupling signal from the transplanted suprachiasmatic nucleus controlling circadian locomotor rhythms. *Nature* 382:810–13

So AY, Bernal TU, Pillsbury ML, Yamamoto KR, Feldman BJ. 2009. Glucocorticoid regulation of the circadian clock modulates glucose homeostasis. *Proc. Natl. Acad. Sci. USA* 106:17582–87

Stephan FK. 2002. The "other" circadian system: food as a Zeitgeber. *J. Biol. Rhythms* 17:284–92

Stephan FK, Swann JM, Sisk CL. 1979. Entrainment of circadian rhythms by feeding schedules in rats with suprachiasmatic lesions. *Behav. Neural. Biol.* 25:545–54

Stokkan KA, Yamazaki S, Tei H, Sakaki Y, Menaker M. 2001. Entrainment of the circadian clock in the liver by feeding. *Science* 291:490–93

Storch KF, Lipan O, Leykin I, Viswanathan N, Davis FC, et al. 2002. Extensive and divergent circadian gene expression in liver and heart. *Nature* 417:78–83

Storch KF, Weitz CJ. 2009. Daily rhythms of food-anticipatory behavioral activity do not require the known circadian clock. *Proc. Natl. Acad. Sci. USA* 106:6808–13

Takahashi JS, Hong HK, Ko CH, McDearmon EL. 2008. The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nat. Rev. Genet.* 9:764–75

Tataroglu O, Davidson AJ, Benvenuto LJ, Menaker M. 2006. The methamphetamine-sensitive circadian oscillator (MASCO) in mice. *J. Biol. Rhythms* 21:185–94

Turek FW, Joshu C, Kohsaka A, Lin E, Ivanova G, et al. 2005. Obesity and metabolic syndrome in circadian *Clock* mutant mice. *Science* 308:1043–45

Ueyama T, Krout KE, Nguyen XV, Karpitskiy V, Kollert A, et al. 1999. Suprachiasmatic nucleus: a central autonomic clock. *Nat. Neurosci.* 2:1051–53

VanderLeest HT, Houben T, Michel S, Deboer T, Albus H, et al. 2007. Seasonal encoding by the circadian pacemaker of the SCN. *Curr. Biol.* 17:468–73

Vitaterna MH, Ko CH, Chang AM, Buhr ED, Fruechte EM, et al. 2006. The mouse *Clock* mutation reduces circadian pacemaker amplitude and enhances efficacy of resetting stimuli and phase-response curve amplitude. *Proc. Natl. Acad. Sci. USA* 103:9327–32

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

Vollmers C, Gill S, DiTacchio L, Pulivarthy SR, Le HD, Panda S. 2009. Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proc. Natl. Acad. Sci. USA* 106:21453–58

Vujovic N, Davidson AJ, Menaker M. 2008. Sympathetic input modulates, but does not determine, phase of peripheral circadian oscillators. *Am. J. Physiol.* 295:R355–60

Webb AB, Angelo N, Huettner JE, Herzog ED. 2009. Intrinsic, nondeterministic circadian rhythm generation in identified mammalian neurons. *Proc. Natl. Acad. Sci. USA* 106:16493–98

Welsh DK, Logothetis DE, Meister M, Reppert SM. 1995. Individual neurons dissociated from rat suprachiasmatic nucleus express independently phased circadian firing rhythms. *Neuron* 14:697–706

Welsh DK, Takahashi JS, Kay SA. 2010. Suprachiasmatic nucleus: cell autonomy and network properties. *Annu. Rev. Physiol.* 72:551–77

Welsh DK, Yoo SH, Liu AC, Takahashi JS, Kay SA. 2004. Bioluminescence imaging of individual fibroblasts reveals persistent, independently phased circadian rhythms of clock gene expression. *Curr. Biol.* 14:2289–95

Yagita K, Tamanini F, van Der Horst GT, Okamura H. 2001. Molecular mechanisms of the biological clock in cultured fibroblasts. *Science* 292:278–81

Yamaguchi S, Isejima H, Matsuo T, Okura R, Yagita K, et al. 2003. Synchronization of cellular clocks in the suprachiasmatic nucleus. *Science* 302:1408–12

Yamamoto T, Nakahata Y, Tanaka M, Yoshida M, Soma H, et al. 2005. Acute physical stress elevates mouse period1 mRNA expression in mouse peripheral tissues via a glucocorticoid-responsive element. *J. Biol. Chem.* 280:42036–43

Yamazaki S, Numano R, Abe M, Hida A, Takahashi R, et al. 2000. Resetting central and peripheral circadian oscillators in transgenic rats. *Science* 288:682–85

Yang X, Downes M, Yu RT, Bookout AL, He W, et al. 2006. Nuclear receptor expression links the circadian clock to metabolism. *Cell* 126:801–10

Yoo SH, Yamazaki S, Lowrey PL, Shimomura K, Ko CH, et al. 2004. PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proc. Natl. Acad. Sci. USA* 101:5339–46

RELATED RESOURCES

Bass J, Takahashi JS. 2010. Circadian integration of metabolism and energetics. *Science* 330:1349–54

Dibner C, Schibler U, Albrecht U. 2010. The mammalian circadian timing system: organization and coordination of central and peripheral clocks. *Annu. Rev. Physiol.* 72:517–49

Do MT, Yau KW. 2010. Intrinsically photosensitive retinal ganglion cells. *Physiol. Rev.* 90:1547–81

Hogenesch JB, Ueda HR. 2011. Understanding systems-level properties: timely stories from the study of clocks. *Nat. Rev. Genet.* 12:407–16

Lowrey PL, Takahashi JS. 2011. Genetics of circadian rhythms in mammalian model organisms. *Adv. Genet.* 74:175–230

Welsh DK, Takahashi JS, Kay SA. 2010. Suprachiasmatic nucleus: cell autonomy and network properties. *Annu. Rev. Physiol.* 72:551–77

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

Annual Review of Neuroscience

Volume 35, 2012

Contents

The Neural Basis of Empathy
Boris C. Bernhardt and Tania Singer 1

Cellular Pathways of Hereditary Spastic Paraplegia
Craig Blackstone 25

Functional Consequences of Mutations in Postsynaptic Scaffolding Proteins and Relevance to Psychiatric Disorders
Jonathan T. Ting, João Peça, and Guoping Feng 49

The Attention System of the Human Brain: 20 Years After
Steven E. Petersen and Michael I. Posner 73

Primary Visual Cortex: Awareness and Blindsight
David A. Leopold 91

Evolution of Synapse Complexity and Diversity
Richard D. Emes and Seth G.N. Grant 111

Social Control of the Brain
Russell D. Fernald 133

Under Pressure: Cellular and Molecular Responses During Glaucoma, a Common Neurodegeneration with Axonopathy
Robert W. Nickells, Gareth R. Howell, Ileana Soto, and Simon W.M. John 153

Early Events in Axon/Dendrite Polarization
Pei-lin Cheng and Mu-ming Poo 181

Mechanisms of Gamma Oscillations
György Buzsáki and Xiao-Jing Wang 203

The Restless Engram: Consolidations Never End
Yadin Dudai 227

The Physiology of the Axon Initial Segment
Kevin J. Bender and Laurence O. Trussell 249

Attractor Dynamics of Spatially Correlated Neural Activity in the Limbic System
James J. Knierim and Kechen Zhang 267

Neural Basis of Reinforcement Learning and Decision Making
Daeyeol Lee, Hyojung Seo, and Min Whan Jung 287

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

Critical-Period Plasticity in the Visual Cortex
Christiaan N. Levelt and Mark Hübener 309

What Is the Brain-Cancer Connection?
Lei Cao and Matthew J. During 331

The Role of Organizers in Patterning the Nervous System
Clemens Kiecker and Andrew Lumsden 347

The Complement System: An Unexpected Role in Synaptic Pruning During Development and Disease
Alexander H. Stephan, Ben A. Barres, and Beth Stevens 369

Brain Plasticity Through the Life Span: Learning to Learn and Action Video Games
Daphne Bavelier, C. Shawn Green, Alexandre Pouget, and Paul Schrater 391

The Pathophysiology of Fragile X (and What It Teaches Us about Synapses)
Asha L. Bhakar, Gül Dölen, and Mark F. Bear 417

Central and Peripheral Circadian Clocks in Mammals
Jennifer A. Mohawk, Carla B. Green, and Joseph S. Takahashi 445

Decision-Related Activity in Sensory Neurons: Correlations Among Neurons and with Behavior
Hendrikje Nienborg, Marlene R. Cohen, and Bruce G. Cumming 463

Compressed Sensing, Sparsity, and Dimensionality in Neuronal Information Processing and Data Analysis
Surya Ganguli and Haim Sompolinsky 485

The Auditory Hair Cell Ribbon Synapse: From Assembly to Function
Saaid Safieddine, Aziz El-Amraoui, and Christine Petit 509

Multiple Functions of Endocannabinoid Signaling in the Brain
István Katona and Tamás F. Freund 529

Circuits for Skilled Reaching and Grasping
Bror Alstermark and Tadashi Isa 559

Indexes

Cumulative Index of Contributing Authors, Volumes 26–35 579

Cumulative Index of Chapter Titles, Volumes 26–35 583

Errata

An online log of corrections to *Annual Review of Neuroscience* articles may be found at http://neuro.annualreviews.org/

Annu. Rev. Neurosci. 2012.35:445-462. Downloaded from www.annualreviews.org
by Stanford University - Main Campus - Green Library on 07/10/12. For personal use only.

ORIGINAL ARTICLE

Childhood Adversities and Adult Psychiatric Disorders in the National Comorbidity Survey Replication I

Associations With First Onset of DSM-IV *Disorders*

Jennifer Greif Green, PhD; Katie A. McLaughlin, PhD; Patricia A. Berglund, MBA; Michael J. Gruber, MS; Nancy A. Sampson, BA; Alan M. Zaslavsky, PhD; Ronald C. Kessler, PhD

Context: Although significant associations of childhood adversities (CAs) with adult mental disorders have been documented consistently in epidemiological surveys, these studies generally have examined only 1 CA per study. Because CAs are highly clustered, this approach results in overestimating the importance of individual CAs. Multivariate CA studies have been based on insufficiently complex models.

Objective: To examine the joint associations of 12 retrospectively reported CAs with the first onset of *DSM-IV* disorders in the National Comorbidity Survey Replication using substantively complex multivariate models.

Design: Cross-sectional community survey with retrospective reports of CAs and lifetime *DSM-IV* disorders.

Setting: Household population in the United States.

Participants: Nationally representative sample of 9282 adults.

Main Outcome Measures: Lifetime prevalences of 20 *DSM-IV* anxiety, mood, disruptive behavior, and substance use disorders assessed using the Composite International Diagnostic Interview.

Results: The CAs studied were highly prevalent and intercorrelated. The CAs in a maladaptive family functioning (MFF) cluster (parental mental illness, substance abuse disorder, and criminality; family violence; physical abuse; sexual abuse; and neglect) were the strongest correlates of disorder onset. The best-fitting model included terms for each type of CA, number of MFF CAs, and number of other CAs. Multiple MFF CAs had significant subadditive associations with disorder onset. Little specificity was found for particular CAs with particular disorders. Associations declined in magnitude with life course stage and number of previous lifetime disorders but increased with length of recall. Simulations suggest that CAs are associated with 44.6% of all childhood-onset disorders and with 25.9% to 32.0% of later-onset disorders.

Conclusions: The fact that associations increased with length of recall raises the possibility of recall bias inflating estimates. Even considering this, the results suggest that CAs have powerful and often subadditive associations with the onset of many types of largely primary mental disorders throughout the life course.

Arch Gen Psychiatry. 2010;67(2):113-123

Author Affiliations:
Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts (Drs Green, McLaughlin, Zaslavsky, and Kessler, Mr Gruber, and Ms Sampson); and Institute for Survey Research, University of Michigan, Ann Arbor (Ms Berglund).

SIGNIFICANT ASSOCIATIONS between retrospectively reported childhood adversities (CAs) and adult illness have been documented in numerous studies.[1,2] The first such studies focused on only a single CA, such as parental death or neglect,[3,4] and 1 mental disorder, most often depression.[5,6] Subsequent studies showed that retrospectively reported CAs are often highly clustered,[7,8] requiring examination of multiple CAs to avoid overestimating associations involving particular CAs.[2,9,10] These studies also found that CAs are often nonspecific in their associations with many different mental disorders,[10-12] making it useful to examine multiple outcomes to avoid overly narrow interpretations.

See also pages 111 and 124

Subsequent studies[13-15] created summary CA scales and documented dose-response relationships with adult outcomes. However, such indices implicitly assumed that each CA has the same effect and that joint effects are additive. These assumptions are almost certainly incorrect.[16] Indeed, a preliminary examination of these assumptions in the National Comorbidity Survey (NCS)[17] showed that some CAs have stronger associations with

©2010 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Alabama-Birmingham User on 07/25/2015

adult outcomes than do others and that joint associations are nonadditive.[10] That study also found that these associations sometimes attenuate with age, a specification generally, but not always,[12,18] ignored in subsequent studies.

The present study builds on these earlier NCS findings by analyzing the CAs assessed in the NCS Replication (NCS-R).[19] Associations between retrospectively reported CAs and mental disorders can be upwardly biased owing to recall failure, nevertheless, it is useful to examine associations based on such retrospective data because they provide upper estimates that avoid the problem of downward bias due to systematic sample attrition in estimates based on long-term prospective data. We examine associations of CAs with the first onset of diverse *DSM-IV* disorders based on several competing multivariate models. A companion study[20] examines associations of CAs with lifetime persistence of the same disorders.

METHODS

SAMPLE

The NCS-R is a face-to-face survey of English-speaking adults performed between February 5, 2001, and April 7, 2003, in a multistage clustered area probability sample of the US household population.[19] The response rate was 70.9%. Recruitment began with a letter and a study fact brochure followed by in-person interviewer visits to explain study aims and procedures before obtaining informed consent. Respondents were paid $50 for participation. Recruitment and consent procedures were approved by the human subjects committees of Harvard Medical School, Boston, Massachusetts, and the University of Michigan, Ann Arbor.

The survey was administered in 2 parts. Part I included a core diagnostic assessment and was administered to all the respondents (n=9282). Part II, which was generally administered on the same occasion as part I, included questions about correlates and additional disorders administered to all part I respondents who met lifetime criteria for any part I disorder plus a probability subsample of other part I respondents (n=5692). The part I sample was weighted to adjust for differential probabilities of selection and intensity of recruitment effort in hard-to-recruit cases. The part II sample, the focus of the present study, was additionally weighted for the lower selection probabilities of part I respondents without a mental disorder. A final weight adjusted the sample to match the 2000 census population on the cross-classification of numerous geographic and sociodemographic variables. All the analyses used these weights. As a result, the sociodemographic characteristics of the weighted part II sample closely match those of the population (eg, 42% female, 71% non-Hispanic white, 24% aged 18-29 years, and 21% ≥60 years old). More detailed information on NCS-R sampling, design, weighting, and sociodemographic distribution is reported elsewhere.[21]

DIAGNOSTIC ASSESSMENT

The NCS-R lifetime diagnoses are based on the World Health Organization Composite International Diagnostic Interview (CIDI),[22] a fully structured, lay-administered interview that generates diagnoses according to the definitions and criteria of the *Internaaational Classification of Diseases, 10th Revision* and the *DSM-IV*. The *DSM-IV* criteria are used herein. The lifetime diagnoses include 4 broad classes of 20 specific disorders: mood disorders (major depressive disorder, dysthymic disorder, bipolar I disorder, bipolar II disorder, and subthreshold bipolar disorder), anxiety disorders (panic disorder, agoraphobia without a history of panic disorder, generalized anxiety disorder, specific phobia, social phobia, posttraumatic stress disorder, and separation anxiety disorder), disruptive behavior disorders (intermittent explosive disorder, attention-deficit/hyperactivity disorder, oppositional-defiant disorder, and conduct disorder), and substance use disorders (alcohol abuse, alcohol dependence with abuse, drug abuse, and drug dependence with abuse). Diagnostic hierarchy rules and organic exclusion rules were used in making diagnoses. The *DSM-IV*/CIDI disorder prevalence estimates in sociodemographic subsamples are reported elsewhere (http://www.hcp.med.harvard.edu/ncs). An NCS clinical reappraisal study[23] found generally good concordance between diagnoses based on the CIDI and those based on blinded clinical reinterviews using the Structured Clinical Interview for *DSM-IV*.[24]

The CIDI assessed age at onset of the disorder retrospectively. Based on evidence that retrospective age-at-onset reports are often erroneous,[25] a special question sequence was used to improve the accuracy of reporting. This began with questions designed to emphasize the importance of accurate responses: "Can you remember your exact age the *very first time* [emphasis in original] when you had [the symptom/the syndrome]?" Respondents who answered "no" were then probed for a bound of uncertainty by asking the earliest age at which they could clearly remember having the disorder. Onset was set at the upper end of the bound of uncertainty. Experimental research[26] has shown that this approach yields more plausible age-at-onset distributions than do standard age-at-onset questions.

CHILDHOOD ADVERSITIES

Twelve dichotomous CAs occurring before age 18 years were assessed in the NCS-R. The selection of CAs was based on a review of the literature. These CAs include 3 types of interpersonal loss (parental death, parental divorce, and other separation from parents or caregivers), 4 types of parental maladjustment (mental illness, substance abuse, criminality, and violence), 3 types of maltreatment (physical abuse, sexual abuse, and neglect), and 2 other CAs (life-threatening childhood physical illness in the respondent and extreme childhood family economic adversity). The measures of parental death, divorce, and other separation (eg, respondent placed in foster care) focus only on the biological parents, not on stepparents or other caregivers. Respondents who were born to a single mother and never experienced any further disruption of this parenting arrangement were coded as not experiencing any parental separation. We did not include information about the number of caregiver disruptions (eg, multiple divorces) or separations (eg, multiple foster care placements) but rather coded respondents dichotomously as having any vs no such disruptions because the rarity of multiple disruptions made estimates of dose-response relationships unstable.

Parental criminality, family economic adversity, and sexual abuse were assessed using a short question series developed for the baseline NCS.[10] Parental criminality was assessed using questions about whether a parent either engaged in criminal activities, such as burglary or selling stolen property, or was ever arrested for criminal activity. Economic adversity was assessed using questions about whether the family received welfare or other government assistance and whether the family often lacked enough money to pay for the basic necessities of living. Sexual abuse was assessed using questions about repeated fondling, attempted rape, and rape. Parental mental

©2010 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Alabama-Birmingham User on 07/25/2015

illness (major depression, generalized anxiety disorder, and prior to panic disorder) and substance abuse were assessed using the Family History Research Diagnostic Criteria interview[27] and its extensions.[28] Family violence and physical abuse of the respondent by parents were assessed using a modified version of the Conflict Tactics Scales.[29] Neglect was assessed using questions used in studies of child welfare about the frequency of not having adequate food, clothing, or medical care; having inadequate supervision; and having to do age-inappropriate chores.[30] Life-threatening physical illness was assessed using a standard chronic conditions checklist.[31]

ANALYSIS METHODS

Tetrachoric factor analysis (promax rotation) was used to examine intercorrelations among CAs. Associations of CAs with lifetime disorders were estimated using discrete-time survival analysis, with person-years as the unit of analysis,[32] controlling for respondent age at interview, sex, race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, and other), and other *DSM-IV*/CIDI disorders with onset before the age at onset of the disorder under investigation and before age 18 years. The controls for early-onset disorders were included to adjust for the associations of CAs with temporally secondary disorders through earlier-onset disorders that affected secondary disorders. Person-years began at age 4 years, the youngest age evaluated for possible disorder onset. Person-years were coded "0" on the dependent variables until the age at onset and "1" at the year of onset and were censored after the year of onset. Several multivariate models were estimated, with each including dummy predictor variables for CAs plus controls. The first model was additive; that is, it included a separate predictor variable for each of the 12 CAs without interaction terms. The second multivariate model included predictor variables for number of CAs without variables for types of CAs experienced. A third model included 12 predictors for type of CA and additional predictors for number of CAs, with the latter starting at exactly 2 rather than 1 because the variable for exactly 1 CA was perfectly predicted by the 12 dummy variables for the individual CAs. A variant of this third model distinguished between 2 types of CAs as described in the "Associations of CAs With the First Onset of *DSM-IV*/CIDI Disorders" subsection. Another variant included interactions between types of CAs and number of CAs. Finally, we considered more complex, inherently nonlinear models, but these did not improve on the fit of the simpler models and are consequently not discussed herein.

The Akaike information criterion[33] was used to select the best multivariate model for the overall data array (ie, the consolidated data file that stacked the 20 separate disorder-specific person-year files and included 19 dummy predictor variables to distinguish among these files, thereby forcing the estimated slopes of disorders on CAs to be constant across disorders). This best-fitting model was then estimated again in subsamples defined by disorder, class of disorder (mood, anxiety, disruptive behavior, and substance use disorders), life course stage, and the conjunction of life course stage with class of disorder. Survival coefficients and their standard errors were exponentiated and are reported as odds ratios (ORs) and 95% confidence intervals, respectively.

The population-attributable risk proportion (PARP) of the outcomes was computed for the best-fitting model. The PARP is the proportion of observed outcomes that would have been prevented in the absence of CAs if the ORs were due to causal effects of CAs.[34] In the more realistic case in which the associations of CAs with outcomes are partly due to common causes, the PARP reflects overall associations. All of the PARPs were calculated using simulation methods to generate individual-level predicted probabilities of the outcome disorders from the coefficients in the best-fitting model with and without coefficients for CAs. The PARP is 1 minus the ratio of the predicted prevalence estimates in the 2 specifications. The PARP for a pooled data set is the average PARP across all disorders included in the calculation based on a constant model across disorders.

All statistical significance tests were evaluated using 2-sided tests ($P < .05$). Because the NCS-R data are clustered and weighted, the design-based Taylor series method[35] implemented in the SUDAAN software system[36] was used to estimate standard errors of ORs.

RESULTS

PREVALENCE AND CO-OCCURRENCE OF CAs

Approximately 53.4% of NCS-R respondents reported having at least 1 CA (**Table 1**). The most common CAs were parental divorce (17.5%), family violence (14.0%), family economic adversity (10.6%), and parental mental illness (10.3%). Multiple CAs were the norm in respondents with each CA, from 51.2% in those with death of a parent to 95.1% in those with parental neglect; there were a mean of 3.2 CAs in respondents with more than 1 CA.

Most tetrachoric correlations between pairs of CAs (94%) are positive. (Detailed results are available on request from the corresponding author.) Negative values are small (range, –0.09 to –0.01). Positive values have a median of 0.11 and an interquartile range (25th-75th percentiles) of 0.04 to 0.19. Factor analysis found 3 meaningful factors (Table 1). Most CAs have significant loadings on the first factor of maladaptive family functioning (MFF) (eg, parental substance abuse, criminality, domestic violence, and abuse and neglect), with factor loadings of 0.32 to 0.67. The second factor represents parental death and other loss with associated economic adversity (factor loadings, 0.50-0.67). The third factor represents parental divorce with associated economic adversity (factor loadings, 0.48-0.83). The CAs in factor 1 are referred to herein as MFF CAs and the remaining CAs as other CAs.

ASSOCIATIONS OF CAs WITH THE FIRST ONSET OF *DSM-IV*/CIDI DISORDERS

In the bivariate models (ie, only 1 CA considered at a time) of the pooled associations of CAs with the first onset of the 20 *DSM-IV*/CIDI disorders, all but 1 CA (parental death) was significant, with ORs of 1.5 to 1.9 for MFF CAs and 1.0 to 1.5 for other CAs (**Table 2**). The ORs are generally smaller in the additive multivariate model, with 8 CAs significant and ORs of 1.0 to 1.4 for MFF CAs and 1.0 to 1.3 for other CAs. The χ^2 test for associations of all CAs is significant ($\chi^2_{12} = 884.5$, $P < .001$), although the ORs are substantively modest. We can reject the hypothesis that the ORs are the same for all CAs ($\chi^2_{11} = 286.6$, $P < .001$).

The multivariate model that considers only number rather than type of CAs shows generally increasing ORs with number of CAs, from 1.3 for 1 CA (compared with

©2010 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Alabama-Birmingham User on 07/25/2015

Table 1. Prevalence of Retrospectively Reported CAs and Promax-Rotated Tetrachoric Factor Loadings (Standardized Regression Coefficients) of CAs Based on a 3-Factor Model (n=5692)[a]

	Individual CAs, % (SE)	Respondents With a Given CA Who Also Had ≥1 Other CA, % (SE)	CAs in Those With >1 CA, Mean No. (SE)	Factor (F) Loadings		
				F1[b]	F2[c]	F3[c]
Interpersonal loss						
Parental death	9.9 (0.5)	51.2 (2.8)	3.1 (0.1)	–0.09	0.67	–0.34
Parental divorce	17.5 (0.8)	63.2 (2.2)	3.4 (0.1)	–0.02	0.00	0.83
Other parental loss	6.7 (0.4)	75.9 (2.7)	3.8 (0.1)	0.07	0.58	0.09
Family maladaptation						
Parental mental illness	10.3 (0.6)	71.7 (2.1)	3.9 (0.1)	0.62	–0.14	–0.20
Parental substance use	8.5 (0.5)	85.5 (1.5)	4.1 (0.1)	0.67	–0.14	–0.01
Parental criminality	7.2 (0.3)	85.3 (1.7)	4.1 (0.1)	0.51	–0.11	0.19
Family violence	14.0 (0.7)	86.6 (1.8)	3.8 (0.1)	0.59	0.10	0.18
Abuse and neglect						
Physical abuse	8.4 (0.5)	87.6 (2.4)	4.3 (0.1)	0.62	0.21	–0.09
Sexual abuse	6.0 (0.2)	72.3 (3.3)	4.1 (0.1)	0.32	0.19	–0.07
Neglect	5.6 (0.4)	95.1 (1.1)	4.6 (0.1)	0.59	0.24	–0.04
Other CAs						
Physical illness	5.8 (0.5)	60.7 (4.2)	3.3 (0.1)	0.14	0.10	–0.17
Economic adversity	10.6 (0.5)	83.4 (2.2)	3.5 (0.1)	–0.01	0.50	0.48
Any adversity	53.4 (1.2)	49.6 (1.1)	3.2 (0.0)			

Abbreviation: CA, childhood adversity.
[a] Correlations among factors: F1-F2: 0.15; F1-F3: 0.24; and F2-F3: 0.07.
[b] Factor 1 refers to maladaptive family functioning CAs.
[c] Factors 2 and 3 are combined and refer to other CAs.

respondents who had no CAs) to highs of 3.4 for 6 CAs and 3.2 for 7 or more CAs. The χ^2 test for the joint associations is significant (χ^2_7=822.0, $P<.001$). The model that includes measures of types of CAs and number of CAs fits the data better than the previous models in terms of Akaike information criterion, as indicated by the types-of-CA measures being significant after controlling for number of CAs (χ^2_{12}=86.9, $P<.001$) and the number of CAs measures being significant after controlling for types (χ^2_6=63.7, $P<.001$). (Detailed results of model fitting are available on request from the author.) The hypothesis that the ORs are the same for all types of CAs can be rejected (χ^2_{11}=60.0, $P<.001$). The MFF CAs consistently have higher ORs than do other CAs. The ORs associated with types are mostly higher than in the additive model, indicating that the additivity assumption led to a downward bias in the estimated associations of individual CAs with the outcome. The reason for this is that the ORs associated with number of CAs in the more complex model are for the most part less than 1.0 and become increasingly smaller as number of CAs increases. This means that although the odds of disorder onset increase with increasing number of CAs, they increase at a significantly decreasing rate.

We also evaluated more complex models but found that they generally did not fit as well as the model with types and number of CAs. One refinement did improve fit, although by distinguishing number of MFF CAs from number of other CAs. The significant subadditive interactions associated with number of CAs are found for MFF CAs (χ^2_6=61.7, $P<.001$) but not for other CAs (χ^2_3=5.2, $P=.16$). (The test had only 3 *df* because no respondents had all 5 MFF CAs.) This was the model used in subsequent disaggregated analyses.

DIFFERENTIAL ASSOCIATIONS BY CLASS OF *DSM-IV*/CIDI DISORDER

Disaggregation shows that CAs are significantly associated with the first onset of each class of disorders (mood, anxiety, disruptive behavior, and substance use). The ORs associated with types of CAs are always associated with increased odds (χ^2_{12}=44.5-193.7, $P<.001$). Those for MFF CAs are more consistently significant (χ^2_7=38.2-115.2, $P<.001$) than are those for other CAs (χ^2_5=7.4-57.5, $P=.19$ to $<.001$) (**Table 3**). The ORs associated with number of CAs are always associated with decreased odds, although they are largely confined to MFF CAs (χ^2_6=19.4-50.9, $P=<.001$ to .004).

Close inspection finds what seems to be meaningful variation in the ORs associated with some MFF CAs, such as parental criminality consistently having its lowest OR and parental substance abuse its highest OR predicting respondent substance use disorders. The more striking pattern, though, is that each MFF CA is significantly associated with each disorder class with rather consistent ORs. The ORs of other CAs are less consistent, with only 25% significant at $P<.05$. Again, there seems to be some meaningful variation, most notably family economic adversity and respondent physical illness associated with anxiety but not mood disorders, but these differences are not statistically significant.

DIFFERENTIAL ASSOCIATIONS BY LIFE COURSE STAGE AND NUMBER OF PREVIOUS DISORDERS

Disaggregation by life course stage (childhood: aged 4-12 years, adolescence: aged 13-19 years, early adulthood: aged 20-29 years, and middle-later adulthood: aged ≥30 years)

©2010 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Alabama-Birmingham User on 07/25/2015

Table 2. Bivariate and Multivariate Associations Between CAs and the Subsequent First Onset of *DSM-IV*/CIDI Disorders (n=5692)[a]

	OR (95% CI)			
	Bivariate[b]	Multivariate (Additive)[c]	Multivariate (No. of CAs)[d]	Multivariate (Interactive)[e]
Maladaptive family functioning CAs				
Parental mental illness	1.7 (1.5-1.8)[f]	1.3 (1.2-1.4)[f]	NA	1.4 (1.3-1.6)[f]
Parental substance abuse	1.8 (1.6-1.9)[f]	1.3 (1.2-1.4)[f]	NA	1.4 (1.2-1.6)[f]
Parental criminality	1.5 (1.4-1.7)[f]	1.0 (1.0-1.2)	NA	1.2 (1.0-1.4)[f]
Family violence	1.8 (1.7-2.0)[f]	1.4 (1.2-1.5)[f]	NA	1.5 (1.3-1.7)[f]
Physical abuse	1.8 (1.7-2.0)[f]	1.2 (1.1-1.4)[f]	NA	1.4 (1.2-1.6)[f]
Sexual abuse	1.8 (1.6-2.0)[f]	1.4 (1.3-1.6)[f]	NA	1.6 (1.4-1.9)[f]
Neglect	1.9 (1.7-2.1)[f]	1.2 (1.0-1.3)[f]	NA	1.4 (1.2-1.6)[f]
χ^2_7	NA	411.2[f]	NA	59.0[f]
Other CAs				
Parental death	1.0 (0.9-1.2)	1.0 (0.9-1.1)	NA	1.1 (0.9-1.2)
Parental divorce	1.1 (1.0-1.3)[f]	1.0 (0.9-1.1)	NA	1.1 (0.9-1.2)
Other parental loss	1.5 (1.4-1.6)[f]	1.2 (1.1-1.3)[f]	NA	1.3 (1.1-1.5)[f]
Physical illness	1.3 (1.2-1.5)[f]	1.3 (1.2-1.4)[f]	NA	1.4 (1.2-1.6)[f]
Economic adversity	1.3 (1.2-1.4)[f]	1.0 (0.9-1.1)	NA	1.1 (1.0-1.3)
χ^2_5	NA	31.7[f]	NA	21.9[f]
χ^2_{12}	NA	884.5[f]	NA	86.9[f]
No. of childhood adversities				
1	NA	NA	1.3 (1.2-1.5)[f]	NA
2	NA	NA	1.8 (1.6-2.0)[f]	1.1 (0.9-1.3)
3	NA	NA	1.9 (1.7-2.2)[f]	0.8 (0.6-1.1)
4	NA	NA	2.4 (2.1-2.7)[f]	0.8 (0.5-1.1)
5	NA	NA	2.8 (2.5-3.1)[f]	0.6 (0.4-1.0)
6	NA	NA	3.4 (2.8-4.1)[f]	0.6 (0.3-1.1)
≥7	NA	NA	3.2 (2.8-3.6)[f]	0.3 (0.2-0.7)[f]
χ^2_7	NA	NA	822.0[f]	χ^2_6=63.7[f]

Abbreviations: CA, childhood adversity; CI, confidence interval; CIDI, Composite International Diagnostic Interview; NA, not applicable; OR, odds ratio.

[a] A separate person-year file was created for each of the 20 disorders, and these 20 files were then stacked. The models were estimated in a discrete-time survival framework with person-year as the unit of analysis using this stacked data set, thereby forcing the slopes to be constant across the 20 disorders. Each model controlled for person-year, age category, sex, 19 dummy variables for the outcome disorder category (ie, for the 20 disorders in the stacked data set), and controls for the previous onset of comorbid conditions that began up to age 17 years. The 5692 respondents had 11 047 disorder onsets, ranging from a low of 101 onsets for bipolar I disorder to a high of 1573 onsets for major depressive disorder. A total of 4 700 780 noncase (ie, not involving 1 of the 11 047 onsets) person-years existed across all disorders in the stacked data set. A 10% stratified probability subsample of these person-years was used as controls, each with a weight of 10 to decrease computation time. No bias in the estimation of ORs is introduced by sampling on the outcome owing to the fact that the sampling fraction cancels out in the estimation of ORs.[37] Estimates of population-attributable risk proportions, though, are biased by subsampling. The weight of 10 (ie, 1 / 10% = 10) was added to correct for this bias. Data on the prevalence of individual CAs and the distribution of the number of CAs separately in person-years with and without onsets of the disorders are available on request. For person-years with an onset, these prevalence estimates range from a low of 9.0% (physical illness) to a high of 28.5% (family violence).

[b] Models were estimated with 1 CA at a time in addition to the controls noted in the previous footnote.

[c] The model was estimated with all 12 CAs in addition to the controls noted in the first footnote.

[d] The model was estimated with dummy predictors for the number of CAs without any information about types of CAs. The same controls used in earlier models were included as well.

[e] The model was estimated with dummy predictors for the number of CAs and information about types of CAs. The same controls used in earlier models were included as well.

[f] Significant at $P<.05$, 2-tailed.

shows that the significant ORs of some, but not all, CAs persist throughout the life course (**Table 4**). The ORs associated with other CAs decline with age, but these declines are generally not statistically significant. The exceptions are significant declines with age in ORs for parental death (χ^2_3=8.1, P=.04), physical abuse (χ^2_3=22.9, $P<.001$), sexual abuse (χ^2_3=40.3, $P<.001$), and physical illness (χ^2_3=13.7, P=.003). The persistence of the OR for other parental loss throughout the life course is striking compared with the OR for parental death being significant only in childhood. More highly disaggregated analyses showed that age-related declines involving sexual abuse were consistent across all disorder classes (although significant only for mood disorders), whereas declines associated with physical abuse, parental death, and physical illness varied by class of disorder. (Detailed results are available on request.)

We also examined differential associations of CAs with the first onset of *DSM-IV*/CIDI disorders as a function of the number of previous lifetime disorders. (Detailed results are available on request from the author.) We found that the ORs associated with most CAs become smaller as the number of previous disorders becomes larger. This means that CAs are more strongly associated with the onset of temporally primary vs secondary disorders. The sign pattern of the associations between types of CAs and onset of disorders remains largely positive (ie, ORs >1.0) when number of previous disorders is 0 (11 of 12 ORs >1.0, 9 of 12 significant at $P<.05$), 1 (7 of 12 ORs >1.0, 0 of 12 significant at $P<.05$), or 2 or more (7 of 12 ORs >1.0, 6 of 12 significant at $P<.05$), but the magnitude of ORs is considerably stronger when number of previous disorders is 0, with median (interquartile range) values of the ORs being higher when number of previous

©2010 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Alabama-Birmingham User on 07/25/2015

Table 3. Multivariate Associations Between CAs and the Subsequent First Onset of *DSM-IV*/CIDI Classes of Disorders Based on a Simple Interactive Model (n=5692)[a]

	OR (95% CI)				
	Mood	Anxiety	Substance Use	Disruptive Behavior[b]	All
MFF CAs					
Parental mental illness	1.8 (1.4-2.3)[c]	1.7 (1.5-2.0)[c]	1.4 (1.0-1.9)[c]	1.8 (1.4-2.3)[c]	1.7 (1.5-1.9)[c]
Parental substance abuse	1.7 (1.4-2.1)[c]	1.4 (1.2-1.6)[c]	2.3 (1.7-3.1)[c]	2.0 (1.5-2.5)[c]	1.7 (1.5-1.9)[c]
Parental criminality	1.3 (1.0-1.7)[c]	1.3 (1.2-1.5)[c]	1.4 (1.1-2.0)[c]	1.7 (1.2-2.3)[c]	1.4 (1.2-1.7)[c]
Family violence	1.4 (1.1-1.8)[c]	1.6 (1.4-1.9)[c]	1.8 (1.4-2.4)[c]	2.0 (1.6-2.6)[c]	1.7 (1.5-2.0)[c]
Physical abuse	1.5 (1.2-1.8)[c]	1.6 (1.3-1.8)[c]	1.6 (1.2-2.1)[c]	2.0 (1.6-2.6)[c]	1.6 (1.4-1.9)[c]
Sexual abuse	2.1 (1.6-2.6)[c]	1.9 (1.6-2.4)[c]	1.7 (1.1-2.4)[c]	1.6 (1.2-2.1)[c]	1.8 (1.5-2.2)[c]
Neglect	1.8 (1.3-2.4)[c]	1.6 (1.3-1.9)[c]	1.8 (1.3-2.5)[c]	1.8 (1.3-2.4)[c]	1.7 (1.4-2.0)[c]
χ^2_7	46.4[ac]	115.2[c]	38.2[c]	53.0[c]	88.0[c]
Other CAs					
Parental death	1.0 (0.8-1.2)	1.2 (1.0-1.4)	1.0 (0.8-1.3)	1.0 (0.8-1.2)	1.1 (0.9-1.2)
Parental divorce	1.0 (0.9-1.2)	1.0 (0.8-1.1)	1.0 (0.8-1.2)	1.1 (0.9-1.3)	1.0 (0.9-1.1)
Other parental loss	1.1 (0.9-1.4)	1.1 (1.0-1.3)	1.5 (1.1-2.0)[c]	1.6 (1.3-2.1)[c]	1.3 (1.1-1.4)[c]
Physical illness	1.2 (1.0-1.5)	1.5 (1.3-1.7)[c]	1.0 (0.8-1.4)	1.5 (1.2-1.9)[c]	1.3 (1.2-1.5)[c]
Economic adversity	1.1 (0.9-1.4)	1.2 (1.0-1.5)[c]	0.9 (0.6-1.2)	1.0 (0.8-1.3)	1.1 (1.0-1.3)
χ^2_5	7.5	57.5[c]	7.4 (0.19)	36.4[c]	35.3[c]
χ^2_{12}	52.5[c]	193.7[c]	44.5[c]	84.2[c]	120.3
No. of MFF CAs					
0-1					
2	0.7 (0.5-1.1)	0.8 (0.6-1.0)[c]	0.6 (0.4-0.9)[c]	0.6 (0.4-0.8)[c]	0.7 (0.6-0.9)[c]
3	0.5 (0.3-0.9)[c]	0.6 (0.5-0.9)[c]	0.4 (0.2-0.7)[c]	0.4 (0.2-0.8)[c]	0.5 (0.4-0.7)[c]
4	0.4 (0.2-0.8)[c]	0.4 (0.3-0.7)[c]	0.2 (0.1-0.5)[c]	0.3 (0.2-0.6)[c]	0.4 (0.2-0.6)[c]
5	0.3 (0.1-0.7)[c]	0.4 (0.2-0.7)[c]	0.2 (0.1-0.6)[c]	0.2 (0.1-0.5)[c]	0.3 (0.1-0.5)[c]
6	0.1 (0.0-0.4)[c]	0.3 (0.1-0.6)[c]	0.1 (0.0-0.3)[c]	0.1 (0.0-0.3)[c]	0.2 (0.1-0.3)[c]
≥7	0.0 (0.0-0.2)[c]	0.2 (0.1-0.3)[c]	0.0 (0.0-0.2)[c]	0.1 (0.0-0.3)[c]	0.1 (0.0-0.2)[c]
χ^2_6	39.8[c]	50.9[c]	19.4[c]	23.9[c]	61.7[c]
No. of other CAs					
0-1					
2	0.7 (0.5-1.0)[c]	0.8 (0.7-1.0)[c]	0.9 (0.7-1.3)	1.0 (0.7-1.2)	0.8 (0.7-1.0)[c]
3	0.8 (0.5-1.3)	0.7 (0.5-1.0)[c]	1.0 (0.6-1.7)	0.9 (0.6-1.4)	0.8 (0.6-1.0)
≥4	0.7 (0.3-1.6)	1.3 (0.6-2.6)	0.6 (0.2-1.8)	0.5 (0.2-1.2)	0.9 (0.6-1.3)
χ^2_3	4.7 (0.20)	13.3 (0.004)	1.2 (075)	3.0 (0.39)	5.2
χ^2_{21}	316.9[c]	1727.0[c]	206.2[c]	465.6[c]	2184.8[c]

Abbreviations: CA, childhood adversity; CI, confidence interval; CIDI, Composite International Diagnostic Interview; MFF, maladaptive family functioning; OR, odds ratio.

[a] See footnote "a" to Table 2 for a description of the data set and overall modeling approach. The model used herein was estimated with predictors for types of CAs and number of CAs (distinguishing number of MFF CAs from number of other CAs) in addition to the controls used in the models described in Table 2. Note that no term was included in the model for having exactly 1 CA. This means that the coefficients for types of CAs can be interpreted as the associations of pure CAs (ie, having 1 and only 1 particular type of CA compared with having none) with onset, whereas the associations with number of CAs represent the extent to which the incremental associations of co-occurring CAs (ie, the added risk of an additional CA in respondents who are otherwise equivalent in terms of the number of other CAs they experienced, controlling for types of other CAs) differ from the associations of pure CAs. The 5692 respondents had 11 047 disorder onsets, including 4545 onsets of an anxiety disorder, 2366 of a substance use disorder, 2357 of a mood disorder, and 1621 of a disruptive behavior disorder. Data on the prevalence of individual CAs and the distribution of the number of CAs separately in person-years with and without onsets of the disorders are available on request. For person-years with an onset, these prevalence estimates range from a low of 7.7% (physical illness associated with onset of a substance use disorder) to a high of 33.5% (family violence associated with onset of a disruptive behavior disorder).

[b] Disruptive behavior disorders are restricted to respondents 44 years and younger at interview.

[c] Significant at $P<.05$, 2-sided test.

disorders is 0 (1.6 [1.2-1.7]) rather than either 1 (1.2 [1.1-1.2]) or 2 or more (1.2 [1.1-1.3]).

POPULATION-LEVEL ASSOCIATIONS OF CAs WITH DISORDER ONSET

We calculated the PARPs associated with CAs based on the best-fitting model. Results show that CAs explain (in a predictive sense) 32.4% of all disorders, 41.2% of disruptive behavior disorders, 32.4% of anxiety disorders, 26.2% of mood disorders, and 21.0% of substance use disorders (**Table 5**). The CAs explain a higher proportion of childhood-onset disorders (44.6%) than adolescent-onset disorders (32.0%) and adult-onset disorders (28.6% and 25.9%). This decline is largely explained by the PARPs for mood disorders decreasing with age from a high of 57.1% for childhood-onset cases to a low of 20.5% for onsets in the age range of 30 years or older. The PARPs also decrease with age for anxiety disorders, but less dramatically than for mood disorders (from 39.5% of childhood-onset cases to 29.8% of onsets in the age range of ≥30 years). The PARPs do not decrease with age, in comparison, for substance use disorders. The number of disruptive behavior disorders that occur for the first time in adulthood is so small that we could not calculate the PARPs for these disorders beyond adolescence.

©2010 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Alabama-Birmingham User on 07/25/2015

Table 4. Multivariate Associations Between CAs and the Subsequent First Onset of *DSM-IV*/CIDI Disorders in 4 Life Course Stages Based on a Simple Interactive Model[a]

	OR (95% CI)			
	Childhood, Aged 4-12 y	Adolescence, Aged 13-19 y	Young Adulthood, Aged 20-29 y	Middle-Later Adulthood, Aged ≥30 y
MFF CAs				
Parental mental illness	1.8 (1.5-2.2)[b]	1.7 (1.4-2.1)[b]	1.5 (1.3-1.9)[b]	1.6 (1.3-1.9)[b]
Parental substance abuse	1.6 (1.3-1.9)[b]	1.8 (1.5-2.2)[b]	1.8 (1.2-2.6)[b]	1.7 (1.3-2.1)[b]
Parental criminality	1.4 (1.1-1.8)[b]	1.6 (1.2-2.0)[b]	1.4 (1.1-1.8)[b]	1.2 (0.9-1.6)
Family violence	1.6 (1.4-1.9)[b]	1.8 (1.4-2.2)[b]	1.8 (1.5-2.3)[b]	1.8 (1.3-2.4)[b]
Physical abuse	1.9 (1.6-2.2)[b]	1.8 (1.4-2.2)[b]	1.5 (1.2-1.9)[b]	1.3 (1.0-1.6)
Sexual abuse	2.3 (1.9-2.7)[b]	1.8 (1.4-2.3)[b]	1.7 (1.3-2.2)[b]	1.4 (1.1-1.9)[b]
Neglect	1.6 (1.3-2.0)[b]	1.8 (1.4-2.3)[b]	1.9 (1.4-2.6)[b]	1.4 (1.0-2.0)[b]
χ^2_7	139.1[b]	49.3[b]	49.5[b]	42.3[b]
Other CAs				
Parental death	1.2 (1.0-1.4)[b]	1.0 (0.8-1.2)	1.0 (0.8-1.4)	1.1 (0.9-1.3)
Parental divorce	1.0 (0.9-1.2)	1.0 (0.9-1.2)	1.0 (0.8-1.2)	1.0 (0.8-1.2)
Other parental loss	1.3 (1.1-1.5)[b]	1.2 (1.0-1.5)[b]	1.2 (0.9-1.6)	1.4 (1.1-1.8)[b]
Physical illness	1.6 (1.4-1.9)[b]	1.2 (1.0-1.4)	1.1 (0.8-1.4)	1.3 (1.0-1.7)
Economic adversity	1.2 (1.0-1.4)[b]	1.0 (0.8-1.2)	1.2 (0.9-1.5)	1.2 (0.9-1.6)
χ^2_5	71.1[b]	12.9[b]	4.2	10.3
χ^2_{12}	342.6[b]	85.9[b]	61.7[b]	81.8[b]
No. of MFF CAs				
0-1				
2	0.8 (0.6-1.0)	0.7 (0.5-0.9)[b]	0.6 (0.4-0.8)[b]	0.7 (0.5-1.1)
3	0.6 (0.4-0.9)[b]	0.5 (0.3-0.7)[b]	0.4 (0.2-0.8)[b]	0.5 (0.3-0.8)[b]
4	0.4 (0.3-0.8)[b]	0.3 (0.1-0.5)[b]	0.3 (0.1-0.5)[b]	0.5 (0.2-0.9)[b]
5	0.3 (0.2-0.7)[b]	0.2 (0.1-0.4)[b]	0.2 (0.1-0.7)[b]	0.4 (0.2-1.1)
6	0.2 (0.1-0.5)[b]	0.1 (0.0-0.2)[b]	0.2 (0.1-0.5)[b]	0.3 (0.1-0.9)[b]
≥7	0.1 (0.0-0.4)[b]	0.0 (0.0-0.2)[b]	0.1 (0.0-0.3)[b]	0.2 (0.0-0.8)[b]
χ^2_6	37.2[b]	47.8[b]	26.4[b]	8.4
No. of other CAs				
0-1				
2	0.9 (0.8-1.1)	0.8 (0.6-1.0)	0.7 (0.5-1.0)	0.8 (0.5-1.2)
3	0.8 (0.6-1.1)	0.8 (0.5-1.2)	0.7 (0.4-1.2)	0.6 (0.3-1.2)
≥4	1.0 (0.4-2.4)	0.5 (0.2-1.2)	0.4 (0.1-1.3)	0.8 (0.3-2.4)
χ^2_3	2.6	4.9	5.4	2.6
χ^2_{21}	1167.8[b]	182.4[b]	472.0[b]	163.0[b]

Abbreviations: CA, childhood adversity; CI, confidence interval; CIDI, Composite International Diagnostic Interview; MFF, maladaptive family functioning; OR, odds ratio.

[a] See footnote "a" to Table 2 for a description of the data set and the overall modeling approach. The model used herein was estimated using predictors for types of CAs and number of CAs (distinguishing number of MFF CAs from number of other CAs) in addition to the controls used in the models described in Table 2. See footnote "a" in Table 3 for a description of the interpretation of the joint effects of type and number of CAs. The 5692 respondents had a total of 11 047 disorder onsets, including 3550 in the age range of 4 to 12 years, 3401 in the age range of 13 to 19 years, 2093 in the age range of 20 to 29 years, and 1845 in the age range of 30 years and older. Data on the prevalence of individual CAs and the distribution of the number of CAs separately in person-years with and without onsets of the disorders are available on request. For person-years with an onset, these prevalence estimates range from a low of 7.7% (physical illness associated with onsets in the age range of 20-29 years) to a high of 31.0% (family violence with onsets in the age range of 4-12 years).

[b] Significant at $P < .05$, 2-sided test.

EFFECTS OF TIME TO RECALL

The use of retrospective data introduces the possibility of recall bias. We investigated this possibility by examining age differences in the reported prevalence of CAs and in the ORs of CAs with disorder onset. (Detailed results are available on request from the author.) Reported death of a parent when the respondent was a child was positively related to age, whereas parental divorce when the respondent was a child was inversely related to age. These patterns are consistent with historical trends. Respondent age was unrelated, in comparison, to reports of other parental loss, neglect, or life-threatening childhood physical illness. Respondent age of 65 years and older was significantly related to low reports of parental mental illness, substance abuse, criminality, family violence, physical abuse, and sexual abuse, whereas age was generally unrelated to these CAs in the age range 18 to 64 years. These patterns could be due to a genuinely low prevalence of some CAs in older respondents, underrepresentation of elderly people with these CAs in the sample (due to early death or differential participation), or underreporting of these CAs in elderly respondents (due to differential recall or differential willingness to report). Although we have no way to know which of these processes are at work, any bias in prevalence estimates is likely conservative in the total sample because of lower reporting in the elderly respondents.

Analysis of age differences in associations at given life course stages found generally good consistency between ORs estimated in the youngest cohorts only (aged 18-29 years at interview) and in the entire sample. Of the

©2010 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Alabama-Birmingham User on 07/25/2015

Table 5. Population-Attributable Risk Proportions (PARPs) of Lifetime *DSM-IV*/CIDI Disorder Types Associated With Childhood Adversities by Life Course Stage[a]

	PARPs				
	Overall	Childhood, Aged 4-12 y	Adolescence, Aged 13-19 y	Early Adulthood, Aged 20-29 y	Middle-Later Adulthood, Aged ≥30 y
Mood	26.2	57.1	30.5	24.7	20.5
Anxiety	32.4	39.5	28.7	31.3	29.8
Substance use	21.0	[b]	26.1	25.6	32.1
Disruptive behavior[c]	41.2	34.4	38.9	[b]	[b]
Any	32.4	44.6	32.0	28.6	25.9

Abbreviation: CIDI, Composite International Diagnostic Interview.

[a] The PARPs were calculated using simulation methods to generate individual-level predicted probabilities of the outcome disorders twice from the coefficients in the best-fitting model: the first time using all the coefficients in the model (probability of the disorder in those exposed to childhood adversities) and the second time assuming that the coefficients associated with the childhood adversities were all zero (probability of the disorder in those unexposed). One minus the ratio of the predicted prevalence estimates in the 2 specifications was then used to calculate PARP. In the pooled data set, the PARP value is the average PARP across all disorders included in the calculation based on a constant model across disorders.

[b] Too few cases available to estimate the PARP.

[c] Disruptive behavior disorders are restricted to respondents 44 years and younger at interview.

48 coefficients for individual CAs (12 CAs associated with disorder onsets in the person-year ranges of 4-12, 13-19, 20-29, and ≥30), 36 were positive and 21 were significant in the youngest cohorts, compared with 41 positive and 31 significant in the total sample. Median (interquartile range) ORs were also similar in the youngest cohorts (1.3 [1.1-1.5]) and the total sample (1.4 [1.2-1.7]). In 8 of 48 cases, the ORs differed significantly for the youngest vs the older cohorts. The OR was significant but was lower in magnitude in the younger (1.2-1.8) vs the older (1.4-1.4) cohorts in 3 of these cases. The OR changed from greater than 1.0 (1.1-1.1) in the older cohorts to less than 1.0 (0.7-0.9) in the younger cohorts in 2 other cases but was nonsignificant in both. The OR was nonsignificant in the youngest cohorts (0.8-1.1) but was significant in older cohorts (1.4-1.7) in the other 3 cases, which involved associations of childhood sexual abuse with disorder onsets in the age ranges of 20 to 29 years and 30 years and older and of parental substance abuse with disorder onsets in the age range of 30 years and older. These findings are not definitive because recall failure could exist even for respondents with the shortest recall intervals, but they nonetheless show that the results are stable across a range of recall times.

COMMENT

Despite these results, this study is limited by the retrospective nature of the data. Methodological research suggests that recall bias can lead to underreporting of CAs,[38] which would be expected to make the estimates of PARPs conservative. However, bias could be anticonservative in estimating ORs if the same respondents who did not report CAs also underreported disorders. A long-term prospective study is needed to resolve these uncertainties. Several such studies[39-42] exist that could be used to evaluate these results, but these studies generally have nontrivial attrition. If this attrition is systematic (ie, respondents with the highest risk of disorders also have the highest attrition), estimates of CA effects could be biased downward. The best way to guard against this possibility is to think of retrospective and prospective studies as bounding the true values of associations (ie, retrospective studies giving upper bound estimates and prospective studies lower bound estimates).

A second study limitation is that the list of CAs, although larger than that in most previous studies, is not exhaustive. We also did not consider the timing, sequencing, persistence, recurrence, or severity of individual CAs. In some cases, such as parental mental illness, there could be complex associations remaining to be discovered that involve the number of ill parents, the number of illnesses, and the persistence and severity of these illnesses. A related limitation is that the analysis of joint CA effects did not include fine-grained evaluation of interactions but focused only on broad interaction patterns. This broad-gauged approach is probably desirable as a first approximation but inevitably misses important subtleties. For example, some research[43-45] suggests that parental divorce is associated with a reduced risk of subsequent psychiatric disorders if it facilitates escape from exposure to maladaptive parenting. Future analyses need to examine such specifications against the backdrop of the broader preliminary patterns found in the present study.

In the context of these limitations, the present results are consistent with those of previous studies in suggesting that most US children are exposed to childhood family adversities that are often clustered.[7,9] Neglect, in particular, almost always appears with other CAs. Even the CAs most likely to be independent co-occur with at least one other CA in most cases. Because of this high co-occurrence, it is critical for future research not to focus on one CA without considering others, because bivariate analyses artificially inflate estimates of individual CA effects.[13] There are implications as well for more subtle analyses. For example, some previous research[11] suggested that childhood neglect exacerbates the predictive effects of other CAs, but the present results raise the possibility that this finding is due to neglect being associated with an especially large number of other un-

©2010 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Alabama-Birmingham User on 07/25/2015

controlled CAs rather than itself creating a high risk of psychiatric disorders.

The present finding that the multivariate structure of the associations between CAs and disorder onset is broadly subadditive has, to our knowledge, never before been examined. This subadditive pattern has important implications for intervention because it means that prevention or amelioration of only a single CA in youths exposed to many CAs is unlikely to have important preventive effects. The finding that this nonadditivity is confined to MFF CAs is reminiscent of the finding in the child maltreatment literature that the most severe CAs tend to be chronic intrafamilial adversities involving the use of physical force.[46] This finding also reinforces the importance of considering CA persistence and severity in future research because the finding that people exposed to many co-occurring MFF CAs have a very high risk of lifetime disorders might be due at least partly to the effects of unmeasured CA persistence and severity.

Despite considerable early theorizing to suggest unique effects of particular CAs on particular mental disorders, such as of childhood parental death on adult depression,[47] we found remarkably little specificity of this sort in the NCS-R data. Most CAs we studied, especially MFF CAs, were associated with all the disorder classes we considered. This pattern was found even in the models that controlled for number of CAs, in which ORs associated with specific CA types can be interpreted under the model as the associations of pure CAs (ie, having a particular 1 and only 1 CA vs none) with disorder onset, thereby removing the confounding effects of CA co-occurrences. We also controlled for comorbid child-adolescent disorders to increase the ability to detect specificities of this sort. Previous studies[48,49] found some evidence of specificity in predicting prevalent cases, but inspection of coefficients in the best-fitting models at the level of disorder class and the level of individual disorders (the latter results are available on request) yielded little evidence of specificity. The obvious implication is that the causal pathways that link CAs to the onset of psychiatric disorders are quite general.

In considering the theme of causal pathways, note that these results do not confirm that CAs have causal effects. An alternative possibility is that unmeasured third variables caused CAs and subsequent mental disorders. Genetic factors are possible confounding variables of this type. This is most obviously true for parental mental illness, which can predict respondent mental illness through genetic pathways unrelated to CAs, but the same might be true for other CAs to the extent that they are indicators of genetic liability. Gene × environment interactions could also be involved to the extent that the people exposed to CAs have an elevated genetic risk of psychiatric disorders and are exposed to stressful experiences related to their CAs that potentiate this genetic liability. Genetically informative designs (eg, twin-family and adoption studies) are needed to evaluate these possibilities rigorously.

Another class of potentially important third variables is respondent behaviors and behavioral predispositions that elicit some CAs, such as abuse, and cause the subsequent onset of respondent mental disorders. Prospective studies that measure these proposed constructs repeatedly would be in the best position to evaluate this possibility. In the ideal case, such studies would have multiple informants to assess reporting bias.

A final noteworthy finding is that the associations of many, but not all, CAs with first onset of *DSM-IV* disorders persist into adulthood. Future research needs to investigate the causal pathways responsible for this specification. Although previous research[15,50] has documented long-term associations of some CAs with adult disorders, these studies almost entirely focused on prevalent cases rather than on first onsets. It is much more striking to document, as we did herein, that CAs continue to be related to first onsets of *DSM-IV* disorders beyond early adulthood. Indeed, the PARPs calculated herein suggest that CAs are associated with more than one-fourth of all new disorders in adulthood. Although several hypotheses could be advanced to explain this finding,[50-52] nothing in these results sheds light on them. The indirect retrospective documentation of long-term multivariate associations is nonetheless important in providing an empirical justification for conducting further analyses to explore such hypotheses to investigate mediators, developmental sequences, and dynamic relationships between CAs and adult-onset disorders.

Future research also needs to distinguish between associations of CAs with disorder onset and disorder persistence. As reported in a companion article,[20] we found a rather different association of CAs with disorder persistence than reported herein with disorder onset. In addition, future research should integrate the kind of broad-based analyses of joint effects presented herein with more focused investigations of specific adversities[53,54] and important adversity clusters.[2,55] Future studies should also examine the moderating effects of early disorders on the associations of CAs with later disorders,[56] a line of study that could be important in focusing clinical attention on preventing the onset of secondary disorders. Finally, future studies should try to identify risk and protective factors in adulthood (eg, personality, social support, and adult stressors) that mediate or modify the relationships of CAs with adult disorders.

Submitted for Publication: February 27, 2009; final revision received August 6, 2009; accepted August 9, 2009.
Correspondence: Ronald C. Kessler, PhD, Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Boston, MA 02115 (kessler@hcp.med.harvard.edu).
Financial Disclosure: Dr Kessler has been a consultant for GlaxoSmithKline Inc, Kaiser Permanente, Pfizer Inc, Sanofi-Aventis, Shire Pharmaceuticals, and Wyeth-Ayerst; has served on advisory boards for Eli Lilly & Co and Wyeth-Ayerst; and has had research support for his epidemiological studies from Bristol-Myers Squibb, Eli Lilly & Co, GlaxoSmithKline, Johnson & Johnson Pharmaceuticals, Ortho-McNeil Pharmaceutical Inc, Pfizer Inc, and Sanofi-Aventis.
Funding/Support: The NCS-R is supported by grant U01-MH60220 from the National Institute of Mental Health (NIMH) with supplemental support from the National Institute on Drug Abuse (NIDA), the Substance Abuse and Mental Health Services Administration, grant 044780 from the Robert Wood Johnson Foundation, and the John

©2010 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Alabama-Birmingham User on 07/25/2015

W. Alden Trust. These activities were supported by grant R01 MH070884 from the NIMH; the John D. and Catherine T. MacArthur Foundation; the Pfizer Foundation; grants R13-MH066849, R01-MH069864, and R01 DA016558 from the US Public Health Service; Fogarty International Research Collaboration Award R03-TW006481 from the Fogarty International Center; the Pan American Health Organization; Eli Lilly & Co; Ortho-McNeil Pharmaceutical Inc; GlaxoSmithKline; and Bristol-Myers Squibb.

Role of the Sponsors: The funding organizations had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

Collaborating NCS-R Investigators: Ronald C. Kessler, PhD (principal investigator, Harvard Medical School); Kathleen Merikangas, PhD (co-principal investigator, NIMH); James Anthony, PhD (Michigan State University); William Eaton, PhD (The Johns Hopkins University); Meyer Glantz, PhD (NIDA); Doreen Koretz, MD (Harvard University); Jane McLeod, PhD (Indiana University); Mark Olfson, MD, MPH (New York State Psychiatric Institute, College of Physicians and Surgeons of Columbia University); Harold Pincus, MD (University of Pittsburgh); Greg Simon, MD, and Michael Von Korff, ScD (Group Health Cooperative); Philip S. Wang, MD, DrPH (NIMH); Kenneth Wells, MD, MPH (UCLA); Elaine Wethington, PhD (Cornell University); and Hans-Ulrich Wittchen, PhD (Max Planck Institute of Psychiatry, Technical University of Dresden).

Disclaimer: The views and opinions expressed herein are those of the authors and should not be construed to represent the views of any of the sponsoring organizations, agencies, or the US government. A complete list of NCS publications and the full text of all NCS-R instruments can be found at http://www.hcp.med.harvard.edu/ncs. Send correspondence to ncs@hcp.med.harvard.edu.

Additional Information: A complete list of World Mental Health (WMH) publications can be found at http://www.hcp.med.harvard.edu/wmh/.

Additional Contributions: The NCS-R is conducted in conjunction with the World Health Organization WMH Survey Initiative. We thank the staff of the World Mental Health (WMH) Data Collection and Data Analysis Coordination Centres for assistance with instrumentation, fieldwork, and consultation on data analysis.

REFERENCES

1. Higgins DJ, McCabe MP. Multiple forms of child abuse and neglect: adult retrospective reports. *Aggress Violent Behav*. 2001;6(6):547-578.
2. Mullen PE, Martin JL, Anderson JC, Romans SE, Herbison GP. The long-term impact of the physical, emotional, and sexual abuse of children: a community study. *Child Abuse Negl*. 1996;20(1):7-21.
3. Fristad MA, Jedel R, Weller RA, Weller EB. Psychosocial functioning in children after the death of a parent. *Am J Psychiatry*. 1993;150(3):511-513.
4. Wark MJ, Kruczek T, Boley A. Emotional neglect and family structure: impact on student functioning. *Child Abuse Negl*. 2003;27(9):1033-1043.
5. Brown GW, Bifulco A, Harris TO. Life events, vulnerability and onset of depression: some refinements. *Br J Psychiatry*. 1987;150:30-42.
6. Comijs HC, Beekman AT, Smit F, Bremmer M, van Tilburg T, Deeg DJ. Childhood adversity, recent life events and depression in late life. *J Affect Disord*. 2007;103(1-3):243-246.
7. Dong M, Anda RF, Felitti VJ, Dube SR, Williamson DF, Thompson TJ, Loo CM, Giles WH. The interrelatedness of multiple forms of childhood abuse, neglect, and household dysfunction. *Child Abuse Negl*. 2004;28(7):771-784.
8. Ney PG, Fung T, Wickett A. The worst combinations of child abuse and neglect. *Child Abuse Negl*. 1994;18(9):705-714.
9. Finkelhor D, Ormrod RK, Turner HA. Poly-victimization: a neglected component in child victimization. *Child Abuse Negl*. 2007;31(1):7-26.
10. Kessler RC, Davis CG, Kendler KS. Childhood adversity and adult psychiatric disorder in the US National Comorbidity Survey. *Psychol Med*. 1997;27(5):1101-1119.
11. Arata CM, Langhinrichsen-Rohling J, Bowers D, O'Brien N. Differential correlates of multi-type maltreatment among urban youth. *Child Abuse Negl*. 2007;31(4):393-415.
12. Collishaw S, Pickles A, Messer J, Rutter M, Shearer C, Maughan B. Resilience to adult psychopathology following childhood maltreatment: evidence from a community sample. *Child Abuse Negl*. 2007;31(3):211-229.
13. Dong M, Anda RF, Felitti VJ, Williamson DF, Dube SR, Brown DW, Giles WH. Childhood residential mobility and multiple health risks during adolescence and adulthood: the hidden role of adverse childhood experiences. *Arch Pediatr Adolesc Med*. 2005;159(12):1104-1110.
14. Dube SR, Anda RF, Felitti VJ, Chapman DP, Williamson DF, Giles WH. Childhood abuse, household dysfunction, and the risk of attempted suicide throughout the life span: findings from the Adverse Childhood Experiences Study. *JAMA*. 2001;286(24):3089-3096.
15. Edwards VJ, Holden GW, Felitti VJ, Anda RF. Relationship between multiple forms of childhood maltreatment and adult mental health in community respondents: results from the adverse childhood experiences study. *Am J Psychiatry*. 2003;160(8):1453-1460.
16. Schilling EA, Aseltine RH, Gore S. The impact of cumulative childhood adversity on young adult mental health: measures, models, and interpretations. *Soc Sci Med*. 2008;66(5):1140-1151.
17. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS. Lifetime and 12-month prevalence of *DSM-III-R* psychiatric disorders in the United States: results from the National Comorbidity Survey. *Arch Gen Psychiatry*. 1994;51(1):8-19.
18. Dube SR, Miller JW, Brown DW, Giles WH, Felitti VJ, Dong M, Anda RF. Adverse childhood experiences and the association with ever using alcohol and initiating alcohol use during adolescence [abstract]. *J Adolesc Health*. 2006;38(4):444.e1-444.e10. http://www.jahonline.org/article/S1054-139X%2805%2900298-3. Accessed May 20, 2009.
19. Kessler RC, Merikangas KR. The National Comorbidity Survey Replication (NCS-R): background and aims. *Int J Methods Psychiatr Res*. 2004;13(2):60-68.
20. McLaughlin KA, Green JG, Gruber MJ, Sampson NA, Zaslavsky AM, Kessler RC. Childhood adversities and adult psychiatric disorders in the National Comorbidity Survey Replication II: associations with persistence of *DSM-IV* disorders. *Arch Gen Psychiatry*. 2010;67(2):124-132.
21. Kessler RC, Berglund P, Chiu WT, Demler O, Heeringa S, Hiripi E, Jin R, Pennell BE, Walters EE, Zaslavsky A, Zheng H. The US National Comorbidity Survey Replication (NCS-R): design and field procedures. *Int J Methods Psychiatr Res*. 2004;13(2):69-92.
22. Kessler RC, Üstün TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res*. 2004;13(2):93-121.
23. Kessler RC, Abelson J, Demler O, Escobar JI, Gibbon M, Guyer ME, Howes MJ, Jin R, Vega WA, Walters EE, Wang P, Zaslavsky A, Zheng H. Clinical calibration of *DSM-IV* diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI). *Int J Methods Psychiatr Res*. 2004;13(2):122-139.
24. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for* DSM-IV *Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP)*. New York: Biometrics Research, New York State Psychiatric Institute; 2002.
25. Simon GE, VonKorff M. Recall of psychiatric history in cross-sectional surveys: implications for epidemiologic research. *Epidemiol Rev*. 1995;17(1):221-227.
26. Knäuper B, Cannell CF, Schwarz N, Bruce ML, Kessler RC. Improving accuracy of major depression age of onset reports in the US National Comorbidity Survey. *Int J Methods Psychiatr Res*. 1999;8(1):39-48.
27. Endicott J, Andreasen N, Spitzer RL. *Family History Research Diagnostic Criteria*. New York: Biometrics Research, New York State Psychiatric Institute; 1978.
28. Kendler KS, Silberg JL, Neale MC, Kessler RC, Heath AC, Eaves LJ. The family history method: whose psychiatric history is measured? *Am J Psychiatry*. 1991;148(11):1501-1504.
29. Straus MA. Measuring intrafamily conflict and violence: the Conflict Tactics (CT) Scales. *J Marriage Fam*. 1979;41(1):75-88.
30. Courtney ME, Piliavin I, Grogan-Kaylor A, Nesmith A. *Foster Youth Transitions to Adulthood: A Longitudinal View of Youth Leaving Care*. Madison, WI: Institute for Research on Poverty; 1998.

©2010 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Alabama-Birmingham User on 07/25/2015

31. Merikangas KR, Ames M, Cui L, Stang PE, Ustun TB, Von Korff M, Kessler RC. The impact of comorbidity of mental and physical conditions on role disability in the US adult household population. *Arch Gen Psychiatry*. 2007;64(10):1180-1188.
32. Willett JB, Singer JD. Investigating onset, cessation, relapse, and recovery: why you should, and how you can, use discrete-time survival analysis to examine event occurrence. *J Consult Clin Psychol*. 1993;61(6):952-965.
33. Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. New York, NY: Springer-Verlag; 2002.
34. Rothman K, Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 1998.
35. Wolter KM. *Introduction to Variance Estimation*. New York, NY: Springer-Verlag; 1985.
36. SUDAAN. *Professional Software for Survey Data Analysis* [computer program]. Version 8.0.1. Research Triangle Park, NC: Research Triangle Institute; 2002.
37. Schlesselman JJ. *Case-Control Studies: Design, Conduct, and Analysis*. New York, NY: Oxford University Press; 1980.
38. Hardt J, Rutter M. Validity of adult retrospective reports of adverse childhood experiences: review of the evidence. *J Child Psychol Psychiatry*. 2004;45(2):260-273.
39. Cohen P, Brown J, Smaile E. Child abuse and neglect and the development of mental disorders in the general population. *Dev Psychopathol*. 2001;13(4):981-999.
40. Fergusson DM, Horwood LJ. The Christchurch Health and Development Study: review of findings on child and adolescent mental health. *Aust N Z J Psychiatry*. 2001;35(3):287-296.
41. Jones GT, Power C, Macfarlane GJ. Adverse events in childhood and chronic widespread pain in adult life: results from the 1958 British Birth Cohort Study. *Pain*. 2009;143(1-2):92-96.
42. Melchior M, Moffitt TE, Milne BJ, Poulton R, Caspi A. Why do children from socioeconomically disadvantaged families suffer from poor health when they reach adulthood? a life-course study. *Am J Epidemiol*. 2007;166(8):966-974.
43. Amato PR. Children's adjustment to divorce: theories, hypotheses, and empirical support. *J Marriage Fam*. 1993;55(1):23-38.
44. Hetherington EM, Stanley-Hagan M. The adjustment of children with divorced parents: a risk and resiliency perspective. *J Child Psychol Psychiatry*. 1999;40(1):129-140.
45. Jaffee SR, Moffitt TE, Caspi A, Taylor A. Life with (or without) father: the benefits of living with two biological parents depend on the father's antisocial behavior. *Child Dev*. 2003;74(1):109-126.
46. Clemmons JC, Walsh K, DiLillo D, Messman-Moore TL. Unique and combined contributions of multiple child abuse types and abuse severity to adult trauma symptomatology. *Child Maltreat*. 2007;12(2):172-181.
47. Tennant C, Bebbington P, Hurry J. Parental death in childhood and risk of adult depressive disorders: a review. *Psychol Med*. 1980;10(2):289-299.
48. McMahon SD, Grant KE, Compas BE, Thurm AE, Ey S. Stress and psychopathology in children and adolescents: is there evidence of specificity? *J Child Psychol Psychiatry*. 2003;44(1):107-133.
49. Shanahan L, Copeland W, Costello EJ, Angold A. Specificity of putative psychosocial risk factors for psychiatric disorders in children and adolescents. *J Child Psychol Psychiatry*. 2008;49(1):34-42.
50. Horwitz AV, Widom CS, McLaughlin J, White HR. The impact of childhood abuse and neglect on adult mental health: a prospective study. *J Health Soc Behav*. 2001;42(2):184-201.
51. Hazel NA, Hammen C, Brennan PA, Najman J. Early childhood adversity and adolescent depression: the mediating role of continued stress. *Psychol Med*. 2008;38(4):581-589.
52. Turner HA, Butler MJ. Direct and indirect effects of childhood adversity on depressive symptoms in young adults. *J Youth Adolesc*. 2003;32(2):89-103.
53. Borges G, Benjet C, Medina-Mora ME, Orozco R, Molnar BE, Nock MK. Traumatic events and suicide-related outcomes among Mexico City adolescents. *J Child Psychol Psychiatry*. 2008;49(6):654-666.
54. Molnar BE, Buka SL, Kessler RC. Child sexual abuse and subsequent psychopathology: results from the National Comorbidity Survey. *Am J Public Health*. 2001;91(5):753-760.
55. Turner HA, Finkelhor D, Ormrod R. The effect of lifetime victimization on the mental health of children and adolescents. *Soc Sci Med*. 2006;62(1):13-27.
56. Espejo EP, Hammen CL, Connolly NP, Brennan PA, Najman JM, Bor W. Stress sensitization and adolescent depressive severity as a function of childhood adversity: a link to anxiety disorders. *J Abnorm Child Psychol*. 2007;35(2):287-299.

©2010 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ by a University of Alabama-Birmingham User on 07/25/2015

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/355058009

Closed-loop neuromodulation in an individual with treatment-resistant depression

Article *in* Nature Medicine · October 2021

DOI: 10.1038/s41591-021-01480-w

CITATIONS
13

READS
612

13 authors, including:



Ankit Khambhati
University of California, San Francisco
68 PUBLICATIONS **1,240** CITATIONS
SEE PROFILE



Leo Sugrue
University of California, San Francisco
20 PUBLICATIONS **2,733** CITATIONS
SEE PROFILE



Kristin Sellers
University of California, San Francisco
49 PUBLICATIONS **842** CITATIONS
SEE PROFILE

All content following this page was uploaded by Kristin Sellers on 11 January 2022.

The user has requested enhancement of the downloaded file.

BRIEF COMMUNICATION
https://doi.org/10.1038/s41591-021-01480-w

nature medicine

Check for updates

Closed-loop neuromodulation in an individual with treatment-resistant depression

Katherine W. Scangos[1,2] ✉, Ankit N. Khambhati[1,3], Patrick M. Daly[1,2], Ghassan S. Makhoul[1,2], Leo P. Sugrue[1,4], Hashem Zamanian[1,2], Tony X. Liu[1,2], Vikram R. Rao[1,5], Kristin K. Sellers[1,3], Heather E. Dawes[1,3], Philip A. Starr[1,3], Andrew D. Krystal[1,2,6] and Edward F. Chang[1,3,6]

Deep brain stimulation is a promising treatment for neuropsychiatric conditions such as major depression. It could be optimized by identifying neural biomarkers that trigger therapy selectively when symptom severity is elevated. We developed an approach that first used multi-day intracranial electrophysiology and focal electrical stimulation to identify a personalized symptom-specific biomarker and a treatment location where stimulation improved symptoms. We then implanted a chronic deep brain sensing and stimulation device and implemented a biomarker-driven closed-loop therapy in an individual with depression. Closed-loop therapy resulted in a rapid and sustained improvement in depression. Future work is required to determine if the results and approach of this *n*-of-1 study generalize to a broader population.

Major depressive disorder (MDD) is a common disorder with high rates of treatment resistance[1]. Deep brain stimulation (DBS) has shown promise as a treatment for refractory MDD but inter-individual response heterogeneity has contributed to inconsistent findings in definitive clinical trials[2–4]. Open-loop approaches deliver fixed, constant stimulation to a single brain structure and have been successful in Parkinson's disease and epilepsy. In MDD, however, evidence that different neural circuits underlie different subsets of MDD symptoms speaks for personalized circuit targeting[5].

There is also need for personalization via temporally controlling stimulation as in closed-loop neuromodulation, where a patient's own physiological activity is used to selectively trigger stimulation only when a pathological state is detected[6]. Because the mood-related effects of neuromodulation exhibit state dependence[7], this temporal specificity may be critical for success in patients with MDD with frequent state changes. Closed-loop stimulation also mitigates concerns for neural adaptation, preserves battery life and reduces side effects. However, closed-loop therapy requires a symptom-specific biomarker that has not previously been identified in MDD. In this study, we report our experience with a patient, where we identified a biomarker of MDD symptoms during a 10-d period of intracranial corticolimbic circuitry mapping. We then successfully implemented the biomarker in implanted closed-loop therapy. This is a demonstration of chronic closed-loop neuromodulation in a psychiatric disorder.

The patient was a 36-year-old woman with childhood onset severe and treatment-resistant MDD (Montgomery–Åsberg Depression Rating Scale (MADRS) = 36 out of 54) unresponsive to multiple antidepressant combinations and electroconvulsive therapy, who participated in a personalized closed-loop neurostimulation trial[7]. We first performed stimulus-response mapping of emotion circuitry employing ten stereoelectroencephalography (SEEG) electrodes implanted bilaterally in the orbitofrontal cortex (OFC), amygdala, hippocampus, ventral capsule/ventral striatum (VC/VS) and subgenual cingulate cortex (SGC). We found dimensionally restrictive clinical responses to stimulation across corticolimbic circuitry aligning with different types of depression symptoms[7]. Neural activity was recorded continuously for 10 d while the patient completed symptom rating scales used to define high and low symptom severity states (Fig. 1a,b).

We then identified SEEG spectral activity features in 6 standard frequency bands that discriminated 15-min segments of resting-state activity coinciding with high and low symptom states ($n = 35$) employing two cross-validated supervised machine learning models at different spatial resolutions. We found that bilateral amygdala gamma power alone was sufficient to detect the high symptom severity state (accuracy: mean = 0.77, s.d. = 0.09; area under the curve (AUC): mean = 0.82, s.d. = 0.11; Fig. 1c and Extended Data Fig. 1a).

We identified right VC/VS as the stimulation site that led to consistent, sustained and dose-dependent improvement of symptoms[7]. We next examined connectivity to determine whether right hemispheric VC/VS and amygdala nodes constituted a structurally and functionally connected subnetwork. We performed evoked potential mapping to generate a global directed network graph[8] (Fig. 1d) and quantified node-relative importance. We found that the amygdala and hippocampus integrated signals from many brain regions (highest weighted indegree) while VC/VS influenced many distant brain regions (highest weighted outdegree; Fig. 1e). Furthermore, the VC/VS-stimulated evoked response was stronger in the amygdala (Fig. 1f) than other network nodes (Extended Data Fig. 1b). Using deterministic tractography we next identified axonal fiber tracts underlying this effective connectivity. Dense right VC/VS-amygdala structural connectivity consisted of the stria terminalis and ansa peduncularis (Fig. 1g and Extended Data Fig. 1c). Finally, VC/VS stimulation at contacts $2^+/3^-$ improved clinical symptoms on 11 out of 13 trials. We examined the change in amygdala gamma power with stimulation with a variety of stimulation paradigms and baseline symptom severities where we had clinical and SEEG data ($n = 5$ trials). We observed a reduction in amygdala gamma power on both trials where symptom severity improved, which was absent on trials when symptom severity did not improve due to alternate

[1]Weill Institute for Neurosciences, University of California, San Francsico, CA, USA. [2]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA. [3]Department of Neurosurgery, University of California, San Francisco, San Francisco, CA, USA. [4]Department of Radiology, University of California, San Francisco, San Francisco, CA, USA. [5]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA. [6]These authors contributed equally: Andrew D. Krystal, Edward F. Chang. ✉e-mail: katherine.scangos@ucsf.edu

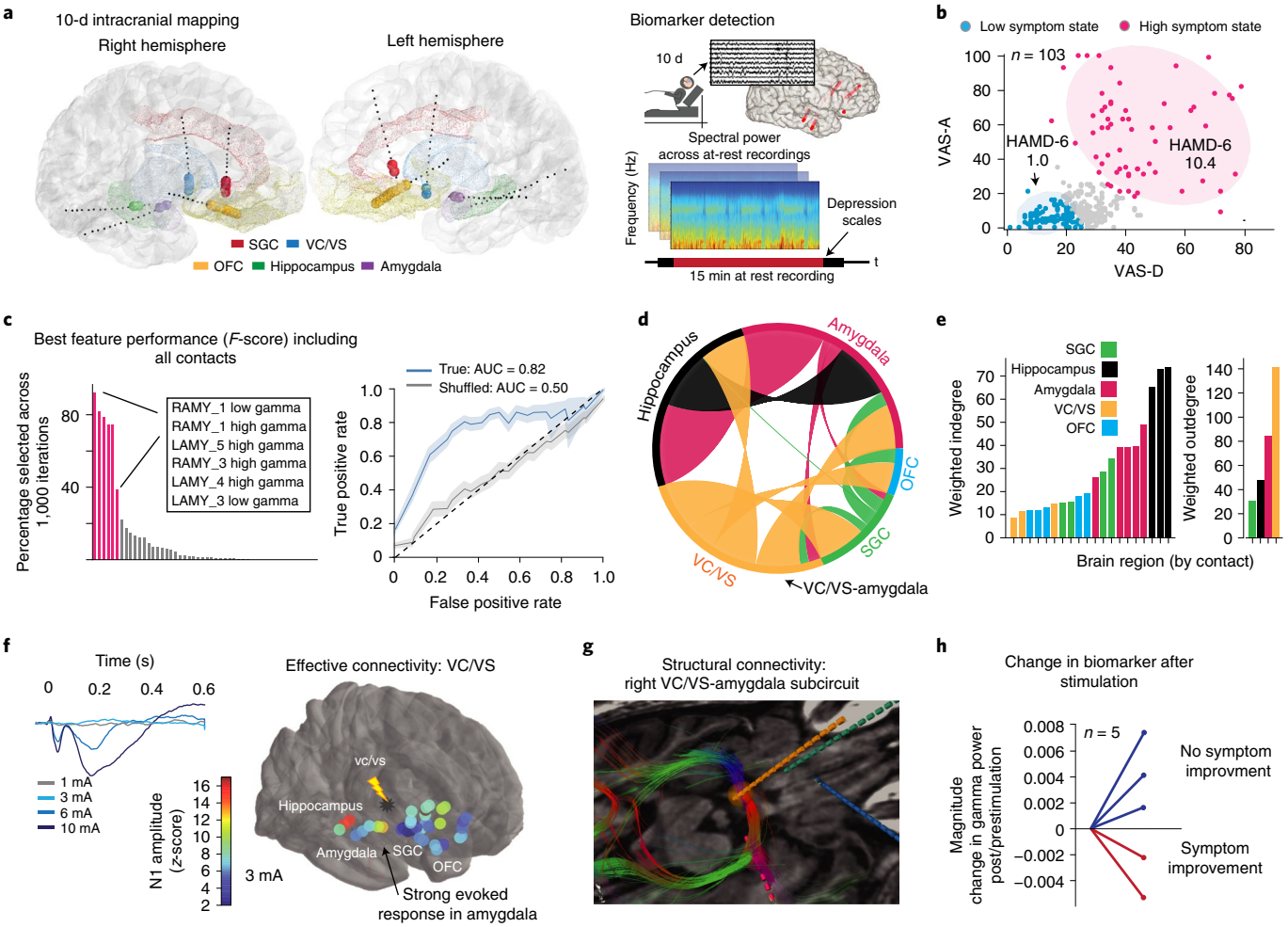


Fig. 1 | Neural biomarker and limbic subnetwork of depression. a, Location of intracranial electrodes and overall approach for biomarker detection. **b**, Clustering clinical reports on anxiety and depression dimensions resulted in two symptom states. Shading has been added to aid visualization. **c**, Top neural features (defined by *F*-score, ANOVA) that discriminated high versus low symptom severity states. Receiver operating characteristic curves reflect the mean ± s.e.m. over $n = 1{,}000$ randomly sampled features for the true (blue) and shuffled models (gray). **d**, Directed network graph of right hemisphere where percentage circumference indicates the strength of connection between any two brain regions. The color indicates the starting location for each set of connections. **e**, Connectivity strength from the network graph was quantified by calculating the weighted indegree and outdegree. The values represent the sum of incoming/outgoing evoked potential waveform N1 (10–50 ms poststimulation) amplitudes, averaged over $n = 20$ repeated pulses. **f**, Example of dose-dependent mean-evoked potentials (left). Summary of evoked potentials across the corticolimbic network due to single pulses in the right VC/VS is shown overlayed on the brain as a heatmap (right). Warmer colors indicate a larger N1 amplitude. **g**, Location of right-sided iEEG leads with fiber tracts (color-coded by orientation) showing structural connections between the amygdala and clinically responsive VC/VS electrodes. Lead location is indicated by color as in **d**,**e**. **h**, Change in gamma biomarker after a period of continuous stimulation for those trials that led to a reduction in symptom severity (red) and those that did not reduce symptom severity (blue).

stimulation paradigms, minimal baseline symptoms or a primary effect on activation over depression (Fig. 1h).

We next implanted the U.S. Food and Drug Administration (FDA)-approved NeuroPace RNS System[9] unilaterally in the right hemisphere (Fig. 2a), with two separate sensing/stimulation four-contact depth leads placed in the amygdala and VC/VS, respectively (Fig. 2b). The NeuroPace RNS System continuously senses neural activity; detection of prespecified patterns of activity are coupled to delivery of electrical stimulation to designated electrodes. Using the NeuroPace RNS System telemetry wand, we wirelessly streamed 10-min trials of continuous amygdala real-time electrocorticography while the patient engaged in naturalistic activities (for example, social media, virtual therapy). We then paired the neural activity with clinical severity ratings (Fig. 2c) and found that the symptom-specific amygdala gamma power replicated findings from the 10-d intracranial mapping. It was highly correlated with both the visual analog scale-depression (VAS-D) ($r^2 = 0.79$, $P = 4.6 \times 10^{-5}$), VAS-anxiety (VAS-A) ($r^2 = 0.53$, $P = 7.5 \times 10^{-3}$) and the 6-item Hamilton Depression Rating Scale (HAMD-6) ($r^2 = 0.68$, $P = 6.5 \times 10^{-4}$; Fig. 2d) and significantly differentiated the high and low symptom severity states (analysis of variance (ANOVA), $F = 6.37$, $P = 0.012$).

The clinical effect of right VC/VS stimulation also replicated results from the intracranial mapping study. Bipolar stimulation at the VC/VS contacts that engaged the stria terminalis and ansa peduncularis (contacts $2^+/3^-$, $3^+/4^-$, 100 Hz, 120 µs, 1–2 mA) led to acute dose-dependent symptom improvement across most trials and the effect was strongest and preferred by the patient (pleasurable/energizing) at the predicted location (contacts $3^+/4^-$) compared to other VC/VS and amygdala contact pairs (Fig. 2e). Furthermore,

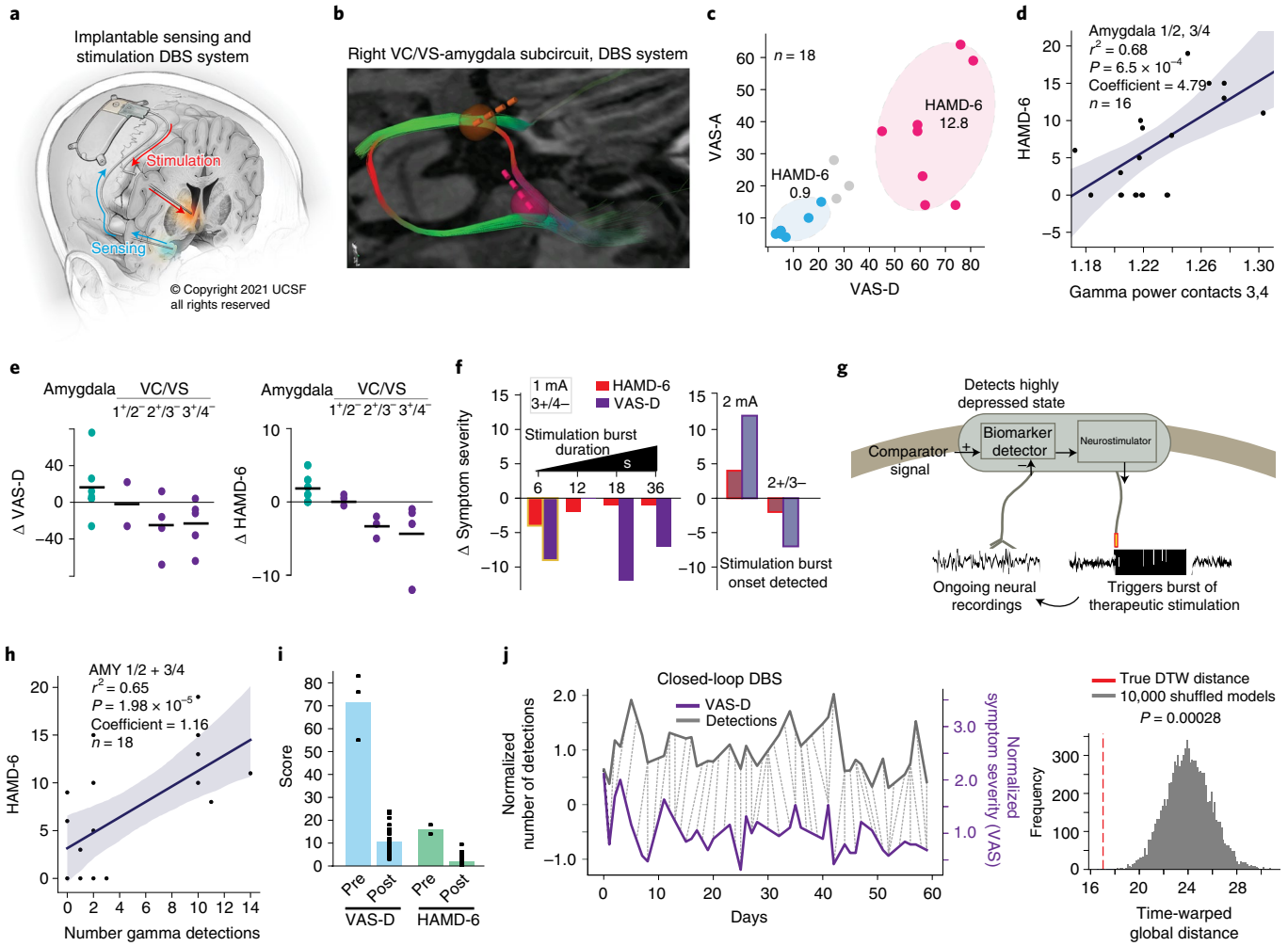


Fig. 2 | Implementation of closed-loop neuromodulation. a, Fully implantable DBS system (illustrated by K. X. Probst). **b**, Reproducibility of targeting showing robust engagement of the stria terminalis and ansa peduncularis. VC/VS lead, yellow; amygdala lead, pink. **c**, Distribution of symptom severity scores in relation to clusters identified in the mapping study. **d**, Positive correlation between gamma power in amygdala contacts 1/2 and 3/4 within the 10-min trials and HAMD-6 score. The linear regression model was evaluated using a two-sided *F*-test and *P* values were adjusted for multiple comparisons. The linear model fit is presented as the mean ± s.e.m. over $n = 16$ symptom ratings. **e**, Reproducibility of clinical effects. Each point represents a stimulation trial ($n = 6, 2, 4, 5$ for VAS-D; $n = 6, 2, 3, 4$ for HAMD) at different bipolar configurations across the right VC/VS and amygdala. **f**, Left: Effect of burst duration on clinical measures. Stimulation parameters: contact $3^+/4^-$, 1 mA, 36 s total stimulation across 15.6 min in 6–36 s intervals. Highlighted condition (6-s burst duration) selected for implementation of closed-loop therapy. Right: Effect of increasing dose (1–2 mA) and changing contacts ($3^+/4^-$ to $2^+/3^-$) on clinical measures. The faded colors indicate that ON/OFF states were detected by the patient (one trial per condition). **g**, Schematic of closed-loop control. **h**, Positive correlation between number of gamma detections by the NeuroPace RNS System within 10-min trials as in **d** and HAMD-6 score. The linear regression model was evaluated using a two-sided *F*-test and *P* values were adjusted for multiple comparisons. The linear model fit is presented as the mean ± s.e.m. over $n = 18$ gamma detections. **i**, Symptom severity scores in the week pre- versus postclosed-loop stimulation onset ($n = 3, 31$ for VAS-D, $n = 2, 30$ for HAMD-6). **j**, Relationship between daily mood ratings (purple) and number of daily biomarker detections (gray). The dotted lines indicate the DTW-computed distance between VAS-D scores and daily biomarker detection numbers (left). Significance was assessed by comparing the DTW distance to that computed from 10,000 randomly scrambled biomarker time series (right).

6 s of intermittent stimulation at 1 mA was clinically effective and below the patient's perceptual threshold (Fig. 2f). Increasing the dose to 2 mA or changing contacts led the patient to perceive stimulation and worsened the clinical response on some trials.

We then implemented closed-loop therapy such that 6 s of stimulation were delivered following the device's automated biomarker detection (Fig. 2g). We found that the number of detections, defined as gamma power crossing a threshold of 0.8% of full amplitude scale within 10-min recording periods was 87% predictive of symptom severity state and highly correlated with VAS-D ($r^2 = 0.59$, $P = 1.2 \times 10^{-4}$), VAS-A ($r^2 = 0.52$, $P = 4.6 \times ^{-4}$) and HAMD-6 ($r^2 = 0.65$, $P = 1.98 \times 10^{-5}$; Fig. 2h). Over the course of 2 months, we found that on average, there were 468 detections distributed across the daytime (s.d. = 206) with minimal nighttime stimulations. We set a stimulation cap of 300 therapies per day (30 min total stimulation) so as not to disturb sleep from evening therapy.

Implementation of closed-loop therapy rapidly improved both symptom severity (measured daily with the HAMD-6/VAS scales) and depression (periodic MADRS). The individual's MADRS score dropped from a 33 before turning treatment ON to a 14 at the first during-treatment assessment carried out after 12 d of stimulation and dropped below 10 (remission) several months later. Similarly, her HAMD-6 and VAS-D scores dropped precipitously the morning after stimulation started (HAMD-6 = 12.0 to 1.0; VAS-D = 77

to 23) and were lower the week after stimulation was turned ON versus the previous week (HAMD-6 = mean 16 (s.d. = 2.82) to 1.6 (s.d. = 1.60), Welch's *t*-test $P = 0.08$; VAS-D = 77.33 (14.57) to 10.48 (5.74), Welch's *t*-test, $P = 0.02$; Fig. 2i).

To evaluate whether our algorithm triggered stimulus delivery linked to patient symptoms, and not randomly, we used dynamic time warping (DTW) to nonlinearly align daily symptom severity (VAS-D) and biomarker detection count time traces over two months and calculated their relative postalignment distance[10]. We found that fluctuation in daily symptoms was significantly associated with the fluctuation in number of device-detected biomarker events ($P = 2.8 \times 10^{-4}$; Fig. 2j). This suggests that our biomarker detection algorithm detects changes in symptom severity significantly better than random chance.

In conclusion, we show the successful development of a personalized biomarker of depression-specific symptoms and implementation of closed-loop therapy for MDD. Success was predicated on a clinical mapping stage before chronic device placement, a strategy that has been utilized in epilepsy to map seizure foci in a personalized manner but has not previously been performed in other neuropsychiatric conditions. During this stage, we developed a comprehensive multimodal framework for selection of sensing and stimulation brain targets (Extended Data Fig. 2). Our approach included personalized stimulus-response mapping, pairing of resting-state signals with clinical symptom measures and identification of functionally and structurally connected subnetworks across the corticolimbic network.

Network mapping and biomarker discovery suggested that VC/VS-amygdala represented an important limbic subnetwork in this patient. Amygdala gamma power reflected depression severity and could be modulated by VC/VS stimulation to decrease clinical symptoms. Previous work showed that these brain regions are important nodes within the mood regulatory circuitry[11]. Bursts of high-frequency amygdala activity correlate with emotion states in humans[12] and VC/VS has been a thoroughly studied DBS target for depression[13] and may be particularly effective for treating anhedonia given its role in the reward pathway. Thus, successful reduction in amygdala gamma may be a requirement for symptom improvement. We validated the amygdala gamma biomarker using the NeuroPace RNS System and identified a set of detection algorithms that enabled us to successfully translate this marker into closed-loop treatment. While traditional DBS is performed bilaterally, we performed right-sided unilateral therapy based on the clinical mapping study and safety considerations. Reasons for observed hemispheric differences are unknown but could be lateralization of emotion regulation or electrode placement within the VC/VS bilaterally.

Our single-participant study cannot evaluate whether the VC/VS-amygdala subcircuit or amygdala biomarker is present in all individuals. Previous literature suggested that the SGC is another promising site for closed-loop stimulation in some individuals[14]. Our methodology intentionally accounts for intersubject variability in MDD symptoms. In contrast to previous DBS studies, a randomized controlled trial using this approach would identify a unique biomarker and treatment site for each individual. It is possible that this personalized approach will overcome limitations evident in previous randomized controlled trials where efficacy was not found when stimulating the VC/VS (or SGC) in all participants[2–4]. In the future, group-level analyses may assess whether certain symptom profiles are predictive of a stimulation target and help define symptom state thresholds that enhance biomarker identification.

We conceptualized MDD as a dynamic process where symptoms arise when a dysfunctional activity state emerges in one or more prefrontal/limbic brain networks subserving mood-related functions. In this model, stimulation has an immediate impact on symptom severity. An acute response to VC/VS stimulation was observed in previous studies in intraoperative or programming periods[15]. Yet, is not always sustained once continuous stimulation is initiated. By delivering intermittent stimulation in a closed-loop manner, we hypothesize that we can repeatedly obtain these acute effects and employ them to treat depression. In line with our model, this patient's depression acutely improved once closed-loop therapy was initiated and the effect has been maintained with sustained closed-loop therapy through the time of writing this article. Both the rapidity and intensity of this clinical response was highly unusual in treatment-resistant depression, where the 1-year remission rate for 'treatment as usual' is approximately 3.5%[16] and symptom relief from DBS can take months to emerge[2].

There are several limitations to this study. First, is the *n*-of-1 sample. It is unknown if the results will generalize. Furthermore, while the patient was blinded to stimulation location and parameters, the clinicians were usually but not always blinded. It is possible that this affected therapeutic response. It is also possible that non-affective sensations elicited during stimulation influenced her affective response. We made efforts to minimize this likelihood by keeping the patient blinded and reducing stimulation intensity if such sensations were observed or reported. Due to device capabilities, we were limited to exploring biomarkers in the power domain in a few frequency bands that could be detected in a time window of a few seconds. New device capabilities that sense continuous neural activity and have the capability to integrate information on the timescale of minutes may improve our ability to incorporate more complex biomarkers into closed-loop control[17]. The degree of network plasticity or drift in the biomarker signal across time is another area for future work. The intent of this study was not to address the efficacy of closed-loop neuromodulation for MDD, which would require a double-blind, randomized controlled, adequately powered study. In this study, we established proof-of-concept for a new powerful treatment approach for neuropsychiatric disorders. The new framework presented in this article could advance biomarker-based neural interfaces and enhance the mechanistic understanding and treatment of a broad range of neuropsychiatric conditions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-021-01480-w.

Received: 3 March 2021; Accepted: 23 July 2021;
Published online: 4 October 2021

References

1. Howland, R. H. Sequenced Treatment Alternatives to Relieve Depression (STAR*D). Part 2: study outcomes. *J. Psychosoc. Nurs. Ment. Health Serv.* **46**, 21–24 (2008).
2. Holtzheimer, P. E. et al. Subcallosal cingulate deep brain stimulation for treatment-resistant depression: a multisite, randomised, sham-controlled trial. *Lancet Psychiatry* **4**, 839–849 (2017).
3. Dougherty, D. D. et al. A randomized sham-controlled trial of deep brain stimulation of the ventral capsule/ventral striatum for chronic treatment-resistant depression. *Biol. Psychiatry* **78**, 240–248 (2015).
4. Bergfeld, I. O. et al. Deep brain stimulation of the ventral anterior limb of the internal capsule for treatment-resistant depression: a randomized clinical trial. *JAMA Psychiatry* **73**, 456–464 (2016).
5. Williams, L. M. Precision psychiatry: a neural circuit taxonomy for depression and anxiety. *Lancet Psychiatry* **3**, 472–480 (2016).
6. Bouthour, W. et al. Biomarkers for closed-loop deep brain stimulation in Parkinson disease and beyond. *Nat. Rev. Neurol.* **15**, 343–352 (2019).
7. Scangos, K. W., Makhoul, G. S., Sugrue, L. P., Chang, E. F. & Krystal, A. D. State-dependent responses to intracranial brain stimulation in a patient with depression. *Nat. Med.* **27**, 229–231 (2021).
8. Keller, C. J. et al. Mapping human brain networks with cortico-cortical evoked potentials. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130528 (2014).

9. Sun, F. T. & Morrell, M. J. The RNS System: responsive cortical stimulation for the treatment of refractory partial epilepsy. *Expert Rev. Med. Devices* **11**, 563–572 (2014).
10. Giorgino, T. Computing and visualizing dynamic time warping alignments in R: the dtw package. *J. Stat. Softw.* **31** (2009).
11. Price, J. L. & Drevets, W. C. Neurocircuitry of mood disorders. *Neuropsychopharmacology* **35**, 192–216 (2010).
12. Kirkby, L. A. et al. An amygdala-hippocampus subnetwork that encodes variation in human mood. *Cell* **175**, 1688–1700.e14 (2018).
13. Malone, D. A. Jr. et al. Deep brain stimulation of the ventral capsule/ventral striatum for treatment-resistant depression. *Biol. Psychiatry* **65**, 267–275 (2009).
14. Mayberg, H. S. et al. Deep brain stimulation for treatment-resistant depression. *Neuron* **45**, 651–660 (2005).
15. Haq, I. U. et al. Smile and laughter induction and intraoperative predictors of response to deep brain stimulation for obsessive-compulsive disorder. *Neuroimage* **54**, S247–S255 (2011).
16. Dunner, D. L. et al. Prospective, long-term, multicenter study of the naturalistic outcomes of patients with treatment-resistant depression. *J. Clin. Psychiatry* **67**, 688–695 (2006).
17. Waltz, E. Green light for deep brain stimulator incorporating neurofeedback. *Nat. Biotechnol.* **38**, 1014–1015 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Participant. In this *n*-of-1 study, the patient was a 36-year-old woman with childhood onset severe, treatment-resistant MDD (MADRS = 36 out of 54) unresponsive to multiple antidepressant combinations and electroconvulsive therapy.

Surgical procedure. The patient gave written informed consent for participation in a clinical trial of closed-loop neuromodulation for treatment-resistant MDD (PReSiDio: NCT04004169), approved by the institutional review board of the University of California, San Francisco (UCSF) and by the FDA. (Please see the Life Sciences Reporting Summary published along with this case report for additional details.) The patient underwent two implant surgeries. In the first, we surgically implanted ten SEEG electrodes (PMT Corporation) within the OFC, amygdala, hippocampus, VC/VS and SGC as described in Scangos et al.[7]. Exploratory intracranial stimulation and recording took place over a 10-d period (October 2019). After 10 d, the electrodes were explanted. In the second implant surgery, we implanted the NeuroPace RNS System with 2 four-contact-depth leads (30-cm lead length, 3.5-mm electrode spacing) in the right VC/VS and amygdala, guided by our findings in the 10-d recording and monitoring (mapping) period. Surgical targeting was planned using the Brainlab iPlan Cranial Software (v3.0) using diffusion tensor imaging (DTI)[18] or coordinate-based targeting[3] in accordance with published work. Computerized tomography was used intraoperatively to confirm electrode placement. No complications of surgery occurred.

Clinical measures. To assess moment-to-moment changes in MDD symptom severity, we used VAS-D[19] and VAS-A and the HAMD-6 of the HAMD-17, which is thought to capture the core symptoms of the full scale and has been used to assess the rapid effects of antidepressants[20–22]. Our symptom assessment strategy included an a priori plan to consider the dimensions of depression that can change in the course of a day as represented in the HAMD-6 subscale of the HAMD-17, which includes: (Q1) sadness; (Q2) guilt; (Q3) apathy; (Q4) fatigue; (Q5) anxiety; and (Q6) energy but focusing only on the dimensions that were possible to meaningfully operationalize in the setting of repeated testing with a VAS (depression, anxiety) and were the smallest number needed to reflect the symptom profile of the patient. Study data were collected and managed using the Research Electronic Data Capture (REDCap, v10.6.19) electronic data capture tools hosted at UCSF[23,24]. REDCap is a secure, Web-based software platform designed to support data capture for research studies, providing (1) an intuitive interface for validated data capture, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages and (4) procedures for data integration and interoperability with external sources. The patient performed her in-lab and at-home surveys using this tool on a study iPad. In the clinical mapping stage, she performed 335 VAS-D, VAS-A and HAMD-6 scales over the 10 d. Scores were distributed across the full range of the patient's natural emotions (VAS-A, mean = 17.5, s.d. = 21.7; VAS-D, mean = 26.3, s.d. = 12.9). These measures were used to define two symptom states (high and low symptom severity). Once the NeuroPace RNS System device was in place, the patient was asked to perform at-home surveys 2–3 times per day.

Signal processing. Intracranial electroencephalogram (iEEG) recordings derived from SEEG electrodes were used to derive measures of activity and connectivity. This is a high-resolution intracranial recording technique[25] and is commonly used in the presurgical evaluation of drug-resistant epilepsy. Stimulation through these electrodes (single pulses[26] or continuous stimulation[27]) was used to perform stimulus-response mapping, assess connectivity and assess the effects of perturbation to MDD networks. Data acquisition of iEEG recordings and evoked potential mapping were performed with a 256-channel Nihon Kohden clinical system and secondary data stream at a sampling rate of 10 kHz. Standard SEEG preprocessing techniques were conducted in Python and involved application of a 2–250 Hz bandpass filter, notch filters at line noise frequency and harmonics, downsampling to 512 Hz and common average referencing across all channels. Signal processing was performed using the continuous wavelet transformation (Morlet, 6-cycles)[28] at 30-s intervals to obtain power in 6 spectral frequency bands (delta = 1–4 Hz, theta = 5–8 Hz, alpha = 9–12 Hz, beta = 13–30 Hz, low gamma = 31–70 Hz, high gamma = 71–150 Hz). Artifacts in evoked potential analysis were removed with an eighth-order Butterworth filter.

Electrode stimulation. Our approach to clinical mapping is discussed elsewhere[7]. Briefly, we tested the clinical effect of a set of stimulation parameters used in the evaluation of patients with epilepsy (1 or 100 Hz, 100 µs, 1–6 mA) through a systematic bipolar stimulation survey and then blinded sham-controlled stimulation studies across 10 d. Where indicated, the brain stimulation configuration is represented by contact number and polarity (for example, $2^{+}/3^{-}$ reflects that contact 2 is anode and contact 3 is cathode). Once the NeuroPace RNS System device was in place, we repeated the bipolar stimulation survey and sham-controlled stimulation using the same clinical scales as outcome measures. The study was single-blinded such that the patient was blinded to stimulation location and parameters. Closed-loop neuromodulation introduces a new stimulation parameter—the burst duration delivered once a depressed state is detected. We tested the effect of burst duration in a systematic manner by delivering increasing durations of stimulation (6, 12, 18, 36 s bursts,100 Hz, 120 µs, 1 mA) while keeping total stimulation time constant over 15- and 30-min time intervals. Our strategy was to identify a set of parameters where intermittent stimulation was clinically effective over a time interval but each burst onset was not detected by the patient to prevent interfering with normal activities and select shorter more frequent bursts over less frequent but longer stimulation periods to increase specificity.

Biomarker discovery. The development of symptom-specific biomarkers in depression have been challenged by the lack of objectively observable clinical symptoms, such as those that exist in movement disorders, and the multidimensional nature of MDD. We addressed this challenge during the mapping trial in several ways. We included a high trial count to address inherent noise in the self-assessment of internal state. We also ensured there was variability across time in the measures by including the presence of mood variability as one of the inclusion criteria of the clinical trial and having the patient perform naturalistic activities in the laboratory setting that induced symptom state variation (for example, recalling life events, watching movies, using social media) during the 10-d period. Then, based on the anxiety and depression dimensions of the VAS scale, we sought to identify states of high and low symptom severity objectively using semisupervised *k*-means clustering. Specifically, to discriminate noisy ratings from clinically relevant ratings, we used cluster tuning and cluster aggregation to discriminate two symptom severity states based on the VAS ratings that best corresponded to the clinically established HAMD-6 scores.

To identify spectral biomarkers related to symptom severity state, we asked the patient to rest for 15-min and then assessed MDD symptom severity (self-reported HAMD-6, VAS scales) about 5 times each day. These 15-min intervals were required to be at least 1 h from stimulation to minimize any potential influence of stimulation on neural activity. Spectral power was calculated in 30-s intervals and averaged across the 15-min recording period to obtain power in 6 frequency bands. We then modeled differences in spectral power across symptom states to identify the spectral power features that were predictive of high symptom severity using cross-validated machine learning models. In the first model, the feature set was defined by spectral power averaged across all contacts within brain regions (10 regions × 6 bands = 60 features). In a second model, the feature set was defined by spectral power for each electrode contact (41 anatomically verified contacts × 6 bands = 246 features). The modeling pipeline consisted of feature selection based on ANOVA *F* value and classification using penalized logistic regression trained on 80% of the data and tested on the remaining 20%. We repeated the training/test schema 1,000 times to ensure stability in AUC score and feature importance histograms. Significance was assessed by comparing the output of our logistic classifier to a null model obtained from randomly permuting class labels 1,000 times.

In closed-loop mode, delivery of therapeutic stimulation is expected to normalize the biomarker that triggers the stimulation. To test the effect of VC/VS stimulation on biomarker activity, we examined the change in gamma power in two 30-s intervals before and after a period of continuous stimulation. We included all types of stimulation trials where we had sufficient iEEG activity and VAS scales. All stimulation was at 100 Hz, 100 µs and 1–3 mA, bipolar contacts $2^{+}3^{-}$. Stimulation was delivered continuously on 4 out of 5 trials. The change was examined in relation to the change in symptom severity due to stimulation.

We validated the biomarker once the NeuroPace RNS System was in place. In the laboratory setting, we streamed continuous neural data (electrocorticography) using the NeuroPace RNS System wand held over the cranially implanted neurostimulator with a flexible metal arm. Once the final RNS detection parameters were identified, we collected 18 trials with 10 min of resting-state neural activity followed by depression measures over the course of 7 d. As in the 10-d mapping period, the patient engaged in a range of naturalistic activities during the recordings. Logistic and linear regression models were then constructed to assess the relationship of amygdala gamma power and symptom severity state or VAS-D and HAMD-6 scores, respectively.

Effective connectivity. Effective connectivity is a functional connectivity measure that describes the directional influence of one brain region on another[29]. Intracranial EEG and intracortical stimulation offer the rare and unique capability to directly probe effective network connectivity through the delivery of brief stimulation pulses across adjacent contacts in one brain region and examination of the electrical effect of perturbation—the evoked potentials—in distant contacts. Evoked potential mapping was performed on day 1. We delivered bipolar single pulse stimulation (1, 3, 6 mA) at 0.5 Hz for 40 s to adjacent contact pairs across a central contact pair in each brain region. Standard methods were used[26] to measure and quantify evoked potentials including calculating the mean across 20 trials, *z*-scoring voltage against 50 ms of prestimulus baseline, identifying N1 and N2 by their peak deflection magnitude within the 10–50 and 50–500 ms time windows, respectively, constructing connectivity matrices from N1 response amplitude for the full set of evoked potentials and filtering on statistically significant evoked potentials to generate a directed network graph. We determined which electrodes exhibited statistically significant evoked potentials by comparing the distribution

of N1 responses due to 20 single pulses to the distribution of prestimulation baseline fluctuations. The baseline score quantified the amount of spontaneous voltage fluctuation in the absence of stimulation and was calculated by finding the peak deflection magnitude during the 50 ms before stimulation, *z*-scored against the mean voltage in the 50–100 ms window before stimulation. We used a one-sided Wilcoxon signed-rank test with Benjamini–Hochberg false discovery rate correction to determine if the N1 response was statistically larger than the baseline spontaneous fluctuation. Compared to a uniform *z*-score threshold, this approach allowed us to account for variable noise between electrodes and identify true evoked potentials with greater sensitivity.

In graph theory, a graph represents the electrophysiology network where nodes are considered to be individual electrodes and the directed link or edge between nodes represents the effect of stimulation of the source node on the amplitude of the neural signal at a distal node[30]. We computed two graph theoretical metrics from the evoked potential connectivity map—indegree and outdegree—to determine the direction and strength of functional connections between nodes and inform the relationship of stimulation and sensing target candidates. A node with high outdegree is a hub of high network influence in that stimulation there highly affects other nodes in the network and could be indicative of a good treatment target[31]. A node with high indegree suggests a region that is influenced by stimulation in other regions and may play a role in sensing modifiable neural signatures of MDD that normalize with stimulation elsewhere in the network. We calculated indegree by taking the weighted sum of significant connections (N1 amplitudes) pointed toward each node and outdegree by the weighted sum of significant connections projecting away from each node for right and left hemispheres and the total brain network. We identified those hubs that had the highest values and were directionally connected.

DTI. DTI can be used to map structural connectivity by identifying putative axonal fiber tracts that might mediate effective connectivity, thereby constraining our model of functional integration across distant brain regions[29]. Engagement of specific white matter tracts or the intersection of several tracts has been shown to improve outcomes in DBS for depression[32,33]. While the exact relationship between structural and functional connectivity is unknown, it has been suggested that structural connectivity properties can be implemented as an informative prior in a Bayesian model of effective connectivity[34]. Diffusion data were acquired using axial DTI high angular resolution diffusion imaging at 3 Tesla with a 32-channel head coil (B value = 2,000 s mm^{-2}, 55 directions). Tractography was performed using deterministic fiber assignment by continuous tracking[35], implemented within the Brainlab Fibertracking (v3.0) software. Tractography was based on 3-mm diameter spherical regions of interest centered on each contact of a candidate stimulation and sensing pair; DTI parameters were the same for all pairs with minimum fractional anisotropy = 0.1, minimum fiber length = 80 mm and maximum angulation = 20 degrees. The resulting fiber tracts are color-coded by orientation (red, left to right; green, anterior-posterior and blue, superior-inferior). Using these parameters, the number of streamlines was used to compare the strength of connectivity for each candidate stimulation-sensing pair (VC/VS stimulation-amygdala sensing, VC/VS stimulation-OFC sensing, SGC stimulation-amygdala sensing and OFC stimulation-amygdala sensing).

Closed-loop control. The NeuroPace RNS System was first placed in a sensing-only configuration where neural biomarkers could be detected but no stimulation was delivered (0 mA). For the first 2.5 months, we found a detection algorithm that could identify the amygdala gamma biomarker, modifying several parameters including detector number, bandpass threshold, window size and count criterion. Our final detector included the following parameters: pattern A1 (amygdala 3/4): minimum-maximum frequency = 28.8–125 Hz, minimum amplitude = 0.8%, window size = 160 ms, count criterion = 10, bandpass threshold = 3, detection analysis window size = 2,048; pattern A2 (amygdala 1/2): minimum-maximum frequency = 28.8–125 Hz, minimum amplitude = 0.8%, window size = 160 ms, count criterion = 10, bandpass threshold = 10, detection analysis window size = 2,048. This means a 'detection' was triggered when gamma power crossed a threshold of 0.8% of full amplitude scale assessed within a 2-s sliding window for amygdala contacts 3/4 or 1/2. While the NeuroPace RNS System does not implement standard frequency decomposition algorithms, the half-wave detector approximates frequency content. In combination with the minimal amplitude settings, the detector tracked gamma activity.

With this detector ON, we examined the number of detections in the 10-min resting-state intervals for the 18 trials utilized for biomarker validation. Logistic and linear regression models were constructed using the number of stimulations as the dependent variable and the symptom state or HAMD-6/VAS-D score as the independent variable. On day 74, the NeuroPace RNS System was turned on in closed-loop mode with stimulation ON (1 mA, 120 µs, 100 Hz). We capped therapy number at 300 per day (30-min total stimulation per day), which was reached at 81.1% of days. On days when the limit was reached, this occurred on average at 19:40 (s.d. = 3.3 h). This limit was selected because stimulation delivered into the evening hours interfered with the patient's sleep. Therapy limit reset time was at 7:00.

To investigate the relationship between daily symptom severity scores (VAS-D) and the time series of the daily number of biomarker detection counts, we used DTW, a method demonstrated to be effective at relating time series exhibiting naturalistic noise[36]. After normalizing symptom severity (VAS-D) and biomarker detection count time traces by their means, we computed the DTW warping function and distance[10]. To prevent the warping function from skipping important symptom severity or biomarker features and cohere with standard practice, we incorporated a Sakoe–Chiba warping window of size three[37]. We tested the significance of the symptom severity and biomarker relationship by comparing our DTW distance to DTW distances computed between symptom severity time series and scrambled biomarker detection time series, where the biomarker time series was obtained by randomly shuffling the observed values 10,000 times.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings in this article are available within the article itself, the source data and within our publicly available GitHub repository. Raw neural signals are freely available from the corresponding author upon request. The data used to produce the results and figures in this paper are available at https://github.com/ScangosLab/closed_loop_mdd. Source data are provided with this paper.

Code availability

The code used to produce the results and figures in this paper is available at https://github.com/ScangosLab/closed_loop_mdd.

References

18. Riva-Posse, P. et al. Defining critical white matter pathways mediating successful subcallosal cingulate deep brain stimulation for treatment-resistant depression. *Biol. Psychiatry* **76**, 963–969 (2014).
19. Campbell, M. C. et al. Mood response to deep brain stimulation of the subthalamic nucleus in Parkinson's disease. *J. Neuropsychiatry Clin. Neurosci.* **24**, 28–36 (2012).
20. Timmerby, N., Andersen, J. H., Søndergaard, S., Østergaard, S. D. & Bech, P. A systematic review of the clinimetric properties of the 6-item version of the Hamilton Depression Rating Scale (HAM-D6). *Psychother. Psychosom.* **86**, 141–149 (2017).
21. Salloum, N. C. et al. Efficacy of intravenous ketamine treatment in anxious versus nonanxious unipolar treatment-resistant depression. *Depress. Anxiety* **36**, 235–243 (2019).
22. Yeung, A. W. K. & Wong, N. S. M. The historical roots of visual analog scale in psychology as revealed by reference publication year spectroscopy. *Front. Hum. Neurosci.* **13**, 86 (2019).
23. Harris, P. A. et al. The REDCap consortium: building an international community of software platform partners. *J. Biomed. Inform.* **95**, 103208 (2019).
24. Harris, P. A. et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42**, 377–381 (2009).
25. Chang, E. F. Towards large-scale, human-based, mesoscopic neurotechnologies. *Neuron* **86**, 68–78 (2015).
26. Matsumoto, R. et al. Functional connectivity in the human language system: a cortico-cortical evoked potential study. *Brain* **127**, 2316–2330 (2004).
27. Rao, V. R. et al. Direct electrical stimulation of lateral orbitofrontal cortex acutely improves mood in individuals with symptoms of depression. *Curr. Biol.* **28**, 3893–3902.e4 (2018).
28. Schiff, S. J., Aldroubi, A., Unser, M. & Sato, S. Fast wavelet transformation of EEG. *Electroencephalogr. Clin. Neurophysiol.* **91**, 442–455 (1994).
29. Friston, K. J. Functional and effective connectivity: a review. *Brain Connect.* **1**, 13–36 (2011).
30. Bassett, D. S. & Sporns, O. Network neuroscience. *Nat. Neurosci.* **20**, 353–364 (2017).
31. Khambhati, A. N. et al. Functional control of electrophysiological network architecture using direct neurostimulation in humans. *Netw. Neurosci.* **3**, 848–877 (2019).
32. Riva-Posse, P. et al. A connectomic approach for subcallosal cingulate deep brain stimulation surgery: prospective targeting in treatment-resistant depression. *Mol. Psychiatry* **23**, 843–849 (2018).
33. Choi, K. S., Riva-Posse, P., Gross, R. E. & Mayberg, H. S. Mapping the "depression switch" during intraoperative testing of subcallosal cingulate deep brain stimulation. *JAMA Neurol.* **72**, 1252–1260 (2015).
34. Sokolov, A. A. et al. Linking structural and effective brain connectivity: structurally informed Parametric Empirical Bayes (si-PEB). *Brain Struct. Funct.* **224**, 205–217 (2019).
35. Mori, S. & van Zijl, P. C. Fiber tracking: principles and strategies—a technical review. *NMR Biomed.* **15**, 468–480 (2002).

36. Rabiner, L. R & Juang, B. H. *Fundamentals of Speech Recognition* (PTR Prentice Hall, 1993).
37. Sakoe, H. & Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust.* **26**, 43–49 (1978).

Acknowledgements

This work was supported by a National Institutes of Health award no. K23NS110962 (K.W.S.), NARSAD Young Investigator grant from the Brain & Behavior Research Foundation (K.W.S.), 1907 Trailblazer Award (K.W.S.) and a Ray and Dagmar Dolby Family Fund through the Department of Psychiatry at UCSF (K.W.S., A.D.K., E.F.C., L.P.S., A.N.K., P.M.S., G.S.M., H.Z., T.X.L., V.R.R., K.K.S. and H.E.D.).

Author contributions

K.W.S., A.N.K., A.D.K. and E.F.C. initiated the work and supervised the study with K.W.S. and A.N.K. contributing equally to the work. H.E.D. and P.A.S. contributed to the conceptualization and evaluation of the study protocol. K.W.S. drafted the manuscript. P.M.D., G.S.M., K.W.S., A.N.K., H.Z., T.X.L., V.R.R., K.K.S. and L.P.S. collected and analyzed the data. A.D.K. and E.F.C. finalized the manuscript. All authors approved the work and take responsibility for its integrity.

Competing interests

A.D.K. consults for Eisai, Evecxia Therapeutics, Ferring Pharmaceuticals, Galderma, Harmony Biosciences, Idorsia, Jazz Pharmaceuticals, Janssen Pharmaceuticals, Merck, Neurocrine Biosciences, Pernix Pharma, Sage Therapeutics, Takeda Pharmaceutical Company, Big Health, Millennium Pharmaceuticals, Otsuka Pharmaceutical and Neurawell Therapeutics. A.D.K. acknowledges support from Janssen Pharmaceuticals, Jazz Pharmaceuticals, Axsome Therapeutics (no. AXS-05-301) and Reveal Biosensors. K.W.S. serves on the advisory board of Nesos. UCSF and E.F.C. have patents related to brain stimulation for the treatment of neuropsychiatric disorders. V.R.R. has served as a paid consultant for NeuroPace but declares no targeted funding from NeuroPace for this study. P.A.S. receives research support from Medtronic. The other authors declare no competing interests.

Additional information

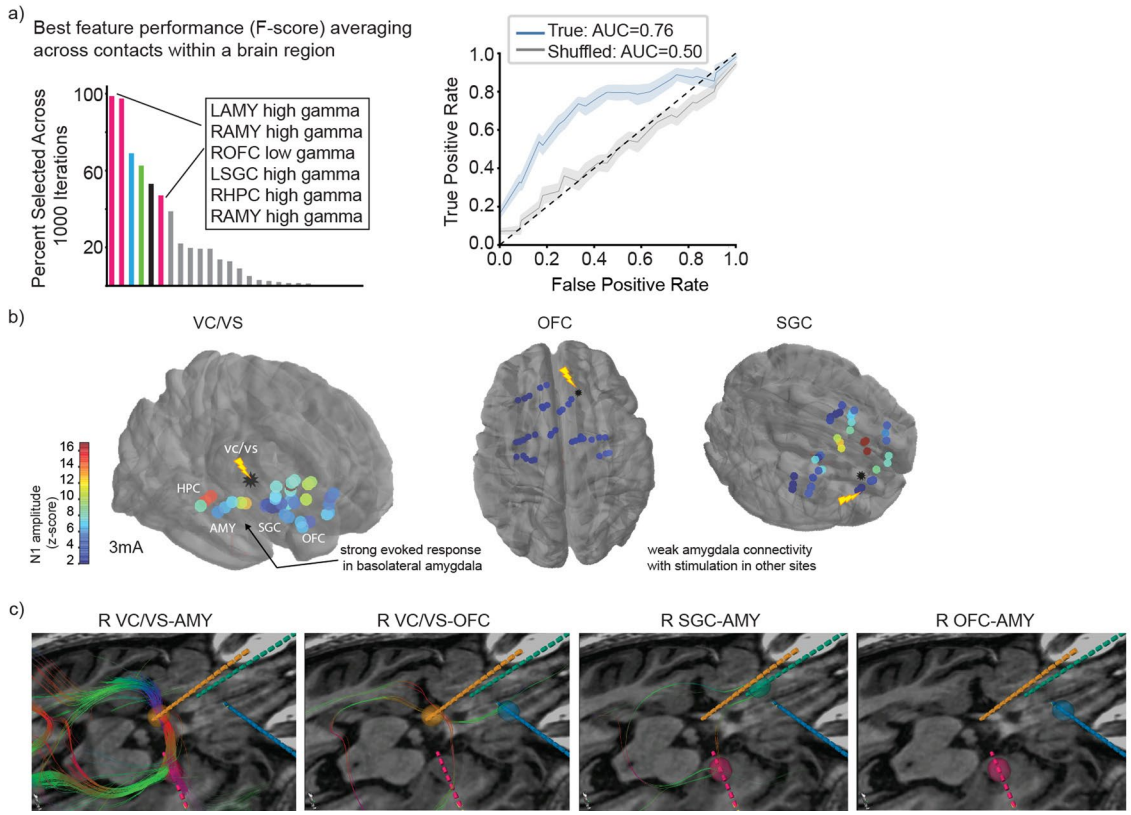
Extended data is available for this paper at https://doi.org/10.1038/s41591-021-01480-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-021-01480-w.

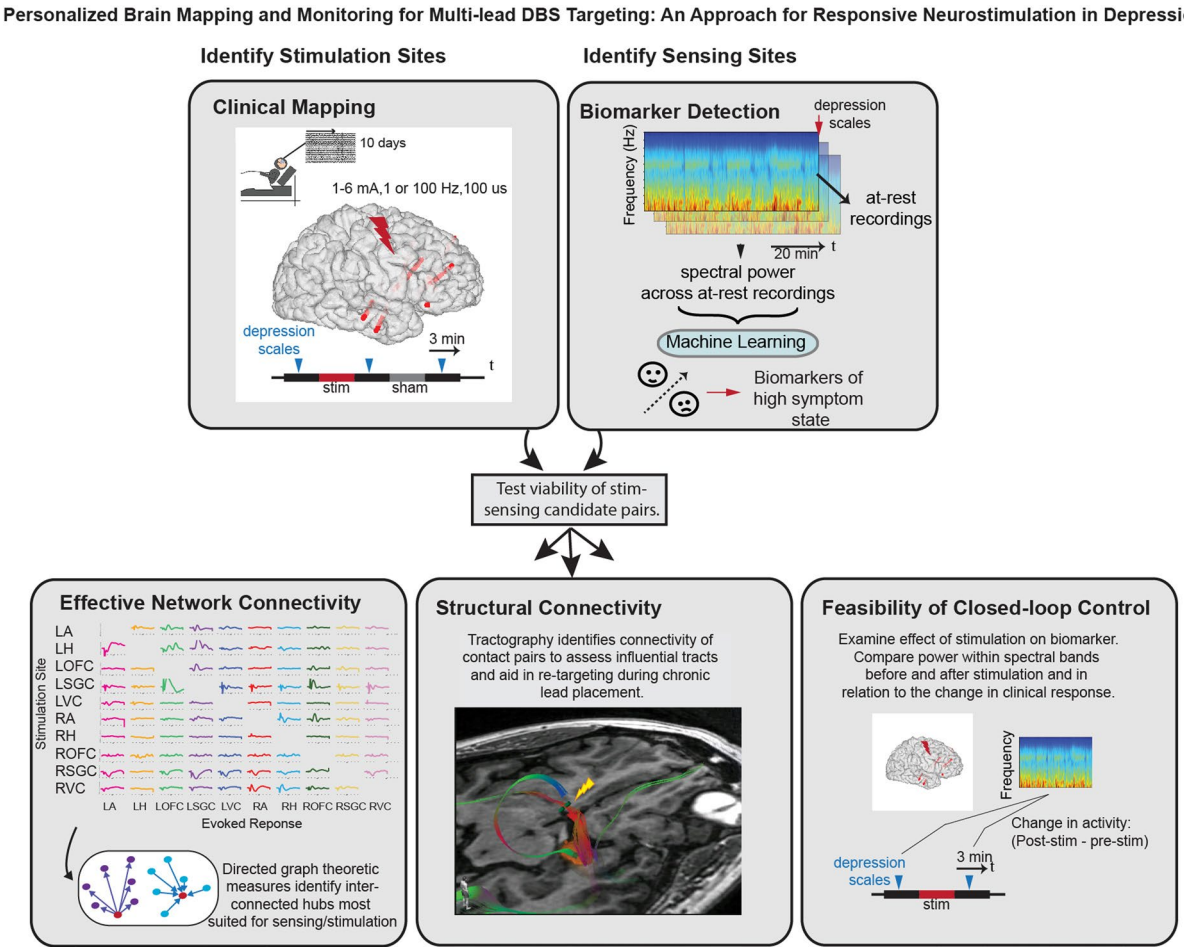
Correspondence and requests for materials should be addressed to Katherine W. Scangos.

Peer review information *Nature Medicine* thanks Ziv Williams and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Jerome Staal was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Network Activity and Connectivity. a. Biomarker identification was performed at two levels of spatial resolution (see also Fig. 1c). In the brain region level model, spectral power was averaged across all contacts within brain regions (60 features). Top neural features (defined by F-score, ANOVA) that discriminated high vs. low symptom severity states are shown. Gamma power in the bilateral AMY, right OFC, left SGC, and right HPC had high state discriminative potential (Accuracy: mean 0.73, std 0.08; AUC: mean 0.76, std 0.10). ROC curves reflect mean ± SEM over n=1000 randomly sampled features for the true model (blue) and the shuffled model (gray). **b**. Evoked potentials (z-scored relative to baseline) across the corticolimbic network due to single pulse stimulation in the right VC/VS, OFC and SGC is shown overlayed on brain as heatmap. Warmer colors indicate a larger N1 amplitude. **c**. Location of right sided sEEG leads targeting AMY (pink), VC/VS (orange), SGC (green), and OFC (blue). Fiber tracts (color coded by orientation) show putative structural connections between candidate pairs of stimulation and sensing contacts (VC/VS-AMY, VC/VS-OFC, SGC-AMY, OFC-AMY) from deterministic tractography using 3 mm ROIs centered on each contact. Tractography parameters were the same for all pairs: minimum FA = 0.1; minimum fiber length = 80 mm; maximum angulation = 20 degrees.



Extended Data Fig. 2 | Overall Approach to Stimulation and Sensing Target Selection. Multimodal method for personalized responsive stimulation multi-lead targeting began with clinical mapping to identify candidate sites where stimulation reliably led to symptom improvement across a range of doses and symptom severity states. Candidate sensing locations were identified by pairing resting state neural activity with symptom severity ratings to identify spectral power biomarkers that correlated with depression. The relationship between candidate stimulation and sensing targets were then tested using three approaches. First, effective network connectivity was assessed by examining the evoked response at nodes across the network due to single pulse stimulation at candidate targets. Second, structural connectivity between candidate contact pairs was assessed using tractography. Influential tracts were identified to help with retargeting during implantation of chronic stimulation device. Finally, the feasibility of closed-loop control was assessed by examining the effect of stimulation in candidate stimulation sites on putative biomarkers identified in candidate sensing locations. This personalized approach enabled us to identify one best stimulation and sensing target pair which were then utilized for the implantation of the RNS System and delivery of closed-loop neurostimulation therapy.

nature research

Corresponding author(s): Katherine Scangos

Last updated by author(s): 7/12/2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	BrainLab iPlan Stereotaxy 3.0, Python 3.5 and 3.7, FSL 6.0, NeuroPace Programmer software v1.8.0.21
Data analysis	The code and data used to produce the results and figures in this paper are available at https://github.com/ScangosLab/closed_loop_mdd

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings in this report are available within the report itself, the source data, and within our publicly available GitHub Repository. Raw neural signals are freely available from the corresponding author upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size is N=1 in this case report.
Data exclusions	No data relevant to the case report was excluded.
Replication	All code was reviewed for accuracy by at least 2 people. All analyses were reproduced successfully. Code to reproduce the figures and analyses are available at: https://github.com/ScangosLab/closed_loop_mdd
Randomization	The data is not randomized as it is a n=1 case-report.
Blinding	The patient was blinded during the sham controlled stimulation studies. Investigators were usually but not always blinded. Blinding was broken during some sessions for safety reasons. We have subsequently added a second physician to each session so that blinding can be maintained.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The patient was a 36 yo woman with severe treatment resistant depression.
Recruitment	The patient contacted our center about participation in a clinical trial of closed-loop DBS (PRESIDIO trial). The patient's physician had heard about the trial and brought it up to the patient in her clinic appointment. She then contacted us about the possiblity of participating. Patients who choose to participate in this study may have a higher tolerance for risk, have more medical knowledge, or have stronger personality disorder traits. They may be more affluent with greater access to or ability to use the internet, or psychiatrists at academic centers who know more about brain stimulation as a treatment option. Due to the inclusion criteria, participants may have depression characterized by greater mood reactivity. As a result, it is possible that closed-loop therapy may be successful only for certain types of depression.
Ethics oversight	The United States Food and Drug Administration (FDA) and the Institutional Review Board (IRB) at University of California, San Francisco approved this protocol. We also work with a neurologist who specializes in ethics of deep brain stimulation.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

View publication stats

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NCT04004169
Study protocol	Please see https://clinicaltrials.gov/ct2/show/NCT04004169. We can provide the study protocol upon reasonable request.
Data collection	Data were collected and analyzed between October 2019 and Jan 2021. Data were collected at the University of California, San Francisco School of Medicine. Specifically, they were collected in a laboratory setting at Langley Porter Psychiatric Hospital and Clinic and in the Epilepsy Monitoring Unit at Moffitt-Long Hospital. Self-report mood ratings were also collected by the patient at home.
Outcomes	The primary and secondary outcomes were pre-defined in our protocol that was submitted to the U.S. Food and Drug Administration and Institutional Review Board at University of California, San Francisco. The measures were logged in ClinicalTrials.gov (https://clinicaltrials.gov/ct2/show/NCT04004169) prior to enrollment of our first subject. We selected the Montgomery Asberg Depression Rating Scale (MADRS) as our primary outcome measure because it is one of the most widely used measures of depression in clinical trials. We additionally utilized visual analog scales and a 6-question version of the Hamilton Depression Rating Scale because these scales have been used previously to assess depression symptoms on a shorter timescale. The MADRS scale was administered by a physician. The remaining scales were administered using RedCap on an iPad in the laboratory or at home. Scores and time of administration were downloaded from RedCap onto a secure server where neural data was also stored.

An *in vivo* correlate of exercise-induced neurogenesis in the adult dentate gyrus

Ana C. Pereira*†, Dan E. Huddleston*†, Adam M. Brickman*†, Alexander A. Sosunov‡, Rene Hen§, Guy M. McKhann‡, Richard Sloan§, Fred H. Gage¶, Truman R. Brown‖, and Scott A. Small*†**

*The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Departments of †Neurology, ‡Neurosurgery, §Psychiatry, and ‖Radiology, Columbia University College of Physicians and Surgeons, New York, NY 10032; and ¶The Salk Institute, La Jolla, CA 92037

Contributed by Fred H. Gage, December 30, 2006 (sent for review November 26, 2006)

With continued debate over the functional significance of adult neurogenesis, identifying an *in vivo* correlate of neurogenesis has become an important goal. Here we rely on the coupling between neurogenesis and angiogenesis and test whether MRI measurements of cerebral blood volume (CBV) provide an imaging correlate of neurogenesis. First, we used an MRI approach to generate CBV maps over time in the hippocampal formation of exercising mice. Among all hippocampal subregions, exercise was found to have a primary effect on dentate gyrus CBV, the only subregion that supports adult neurogenesis. Moreover, exercise-induced increases in dentate gyrus CBV were found to correlate with postmortem measurements of neurogenesis. Second, using similar MRI technologies, we generated CBV maps over time in the hippocampal formation of exercising humans. As in mice, exercise was found to have a primary effect on dentate gyrus CBV, and the CBV changes were found to selectively correlate with cardiopulmonary and cognitive function. Taken together, these findings show that dentate gyrus CBV provides an imaging correlate of exercise-induced neurogenesis and that exercise differentially targets the dentate gyrus, a hippocampal subregion important for memory and implicated in cognitive aging.

hippocampus | *in vivo* imaging | cerebral blood volume | angiogenesis

The hippocampal formation, a brain circuit made up of separate but interconnected subregions (1), is vital for memory function (2) and is targeted by the aging process (3). The dentate gyrus is the only hippocampal subregion that supports neurogenesis in the adult brain (4–6). Nevertheless, because neurogenesis can only be assessed in postmortem tissue, its functional significance remains undetermined. With this limitation in mind, we have explored different imaging approaches applicable to rodents and humans that might provide an *in vivo* correlate of neurogenesis.

Although imaging radioligands designed to bind newly dividing cells is an attractive approach, positron emission tomography imaging suffers inherently poor resolution and cannot visualize the dentate gyrus. Additionally, radiolabeling newborn cells introduces potential safety concerns. For these reasons, we have focused on MRI technologies instead. Notably, a coupling has been established between neurogenesis and angiogenesis (7, 8). The process of angiogenesis, in turn, gradually gives rise to the formation of new blood vessels, increasing regional microvascular density (9–12). Importantly, vascular density can be measured *in vivo* with imaging techniques that map regional blood volume. Numerous studies have established a tight relationship between angiogenesis and regional blood volume (13–17), including in the brain where regional angiogenesis is coupled to regional cerebral blood volume (CBV) (18–26).

Because CBV can be measured with MRI, we hypothesized that a regionally selective increase in hippocampal CBV might provide an imaging correlate of neurogenesis. This hypothesis was tested in mice in which *in vivo* imaging and postmortem analysis can be performed on the same subjects. We used an MRI approach that can generate hippocampal CBV maps repeatedly and safely over time in mice while simultaneously assessing multiple hippocampal subregions (27). We exploited the capabilities of this approach to establish how hippocampal CBV maps change during physical exercise, a potent stimulant of dentate gyrus neurogenesis (28), and tested whether CBV changes correlate with postmortem measurements of neurogenesis. Once these findings were established in mice, we were interested in determining how exercise affects the hippocampal CBV maps of humans. To accomplish this goal, we first optimized an MRI approach (29) previously shown to be capable of generating hippocampal CBV maps in non-human primates (30) and then charted exercise-induced CBV changes in the human hippocampal formation.

Results

Exercise Selectively Increases Dentate Gyrus CBV in Mice and Correlates with Neurogenesis. In designing our experimental protocol, we were guided by the observation that angiogenesis-induced sprouting of new blood vessels progresses through different stages, forming gradually over time (9). Accordingly, mice were allowed to exercise for 2 weeks, the period during which neurogenesis reaches its maximum increase, and bromodeoxyuridine (BrdU), a marker of newly born cells, was injected daily during the second week. To capture the predicted delayed effect in CBV, mice were kept alive for 4 more weeks, then killed and processed for BrdU labeling. Hippocampal CBV maps (Fig. 1*b*) assessing the entorhinal cortex, the dentate gyrus, and the CA3 and CA1 subfields were generated four times over the 6-week experiment: at preexercise baseline and at weeks 2, 4, and 6. A control group following the identical protocol but without exercise was imaged in parallel.

A repeated-measure ANOVA was used to analyze the imaging data set (Fig. 1*a*). A group–time interaction was found only for the dentate gyrus, showing that exercise was associated with a selective increase in dentate gyrus CBV ($F = 5.0$, $P = 0.034$). As shown by simple contrasts, the effect was driven by a maximum increase that emerged 2 weeks after the cessation of exercise, from week 2 to week 4 ($F = 5.9$, $P = 0.021$) (Fig. 1*a*). The entorhinal cortex was the only other hippocampal subregion whose CBV increased appreciably over time, although not achieving statistical significance (Fig. 1*a*). Although exercise might potentially affect CBV by increasing metabolism and cerebral blood flow, previous studies (31, 32) have shown that exercise-induced changes in metabolism should manifest during, not after, the exercise regimen. Thus, the observed spatiotemporal profile with which CBV emerged fits better with a model of exercise-induced angiogenesis in the dentate gyrus.

In agreement with previous studies (33), the exercise group was found to have greater BrdU labeling compared with the nonexercise group ($F = 9.8$, $P = 0.004$) (Fig. 2*a*). More than 90% of

Author contributions: G.M.M., R.S., F.H.G., and S.A.S. designed research; A.C.P., D.E.H., A.M.B., and A.A.S. performed research; R.H. and T.R.B. contributed new reagents/analytic tools; and A.C.P., F.H.G., and S.A.S. wrote the paper.

The authors declare no conflict of interest.

Abbreviation: CBV, cerebral blood volume.

**To whom correspondence should be addressed. E-mail: sas68@columbia.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0611721104/DC1.

© 2007 by The National Academy of Sciences of the USA

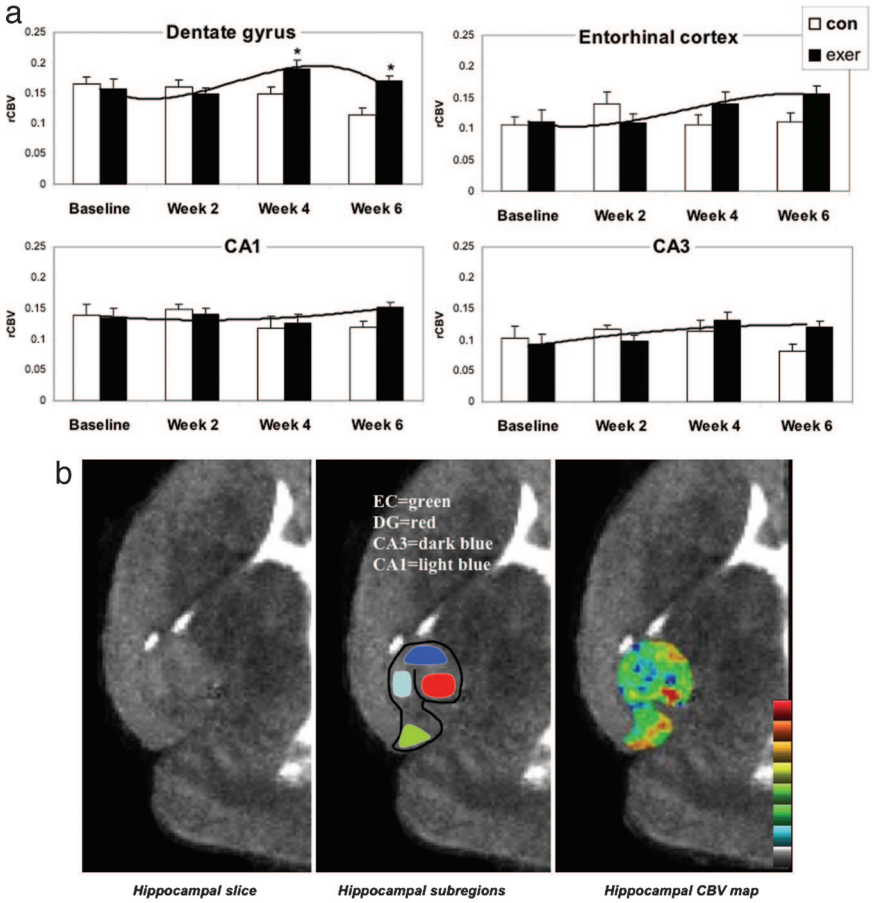


Fig. 1. Exercise selectively increases dentate gyrus CBV in mice. (*a*) Exercise had a selective effect on dentate gyrus CBV. Bar graphs show the mean relative CBV (rCBV) values for each hippocampal subregion in the exercise group (filled bars) and the nonexercise group (open bars) over the 6-week study. The dentate gyrus was the only hippocampal subregion that showed a significant exercise effect, with CBV peaking at week 4, whereas the entorhinal cortex showed a nonsignificant increase in CBV. (*b*) An individual example. (*Left*) High-resolution MRI slice that visualizes the external morphology and internal architecture of the hippocampal formation. (*Center*) Parcellation of the hippocampal subregions (green, entorhinal cortex; red, dentate gyrus; dark blue, CA3 subfield; light blue, CA1 subfield). (*Right*) Hippocampal CBV map (warmer colors reflect higher CBV).

BrdU-positive cells colabeled for NeuN, a neuron-specific marker (Fig. 2*a*). To examine the relationship between neurogenesis and CBV, the repeated-measures model was again used and included BrdU as a covariate. A significant time-by-BrdU interaction was observed only for dentate gyrus CBV ($F = 3.3$, $P = 0.039$), driven primarily by changes from week 2 to week 4 ($F = 8.8$, $P = 0.006$). As shown by a direct analysis, this effect reflected a positive correlation between BrdU and changes in CBV from week 2 to week 4 ($\beta = 0.58$, $P = 0.001$) (Fig. 2*b*). Of note, when BrdU was included as a covariate in the ANOVA, the group–time effect observed in the dentate gyrus was no longer significant, confirming that neurogenesis accounted for the exercise effect on CBV. Visual inspection of the relationship between changes in dentate gyrus CBV and BrdU (Fig. 2*b*) suggested that a quadratic vs. a linear model better characterized the relationship, which was confirmed by curve estimation analysis (linear model, $R^2 = 0.34$ and $P = 0.001$; quadratic model, $R^2 = 0.59$ and $P < 0.0001$). Thus, the association between changes in dentate gyrus CBV and BrdU exists primarily when CBV increases with exercise (Fig. 2*b*). Finally, to test whether neurogenesis was required for the observed increases in dentate gyrus CBV we relied on a previously described x-irradiation approach that selectively blocks dentate gyrus neurogenesis (55). Four mice received x-irradiation, and three mice were sham controls. After a 3-month recovery period, all mice were imaged by using the exercise protocol. A repeated-measure ANOVA applied to the CBV data set revealed a group (x-ray vs. sham)-by-time interaction only for the dentate gyrus ($F = 7.6$, $P = 0.04$), showing that irradiation blocked the exercise-induced increases in CBV [supporting information (SI) Fig. 5].

Exercise Selectively Increases Dentate Gyrus CBV in Humans and Correlates with Aerobic Fitness and Cognition. Once we established that dentate gyrus CBV provides a correlate of exercise-induced neurogenesis, we were interested in testing whether this effect is observed in exercising humans. CBV maps of the human hippocampal formation were generated by using our previously reported MRI approach, specifically tailored for imaging the primate hippocampal formation (30). Eleven healthy subjects (mean age = 33, ranging from 21–45 years; two males and nine females) participated in the study, completing a 3-month aerobic exercise regimen. Hippocampal CBV maps were generated before and after exercise. CBV values were reliably measured for all hippocampal subregions, except the CA3 subregion (Fig. 3*b*). Compared with experimental animals, in humans it is impossible to control the interindividual differences in physical activity performed during daily life. Therefore, before and after exercise, we measured the maximum volume of oxygen consumption (VO_2max), the gold standard measure of exercise-associated aerobic fitness (34, 35) to quantify individual differences in degree of exercise. Cognitive performance was assessed by using a modified Rey Auditory Verbal Learning Test (36), whose design allows cognition to be tested across different learning trials and during delayed recall, recognition, and source

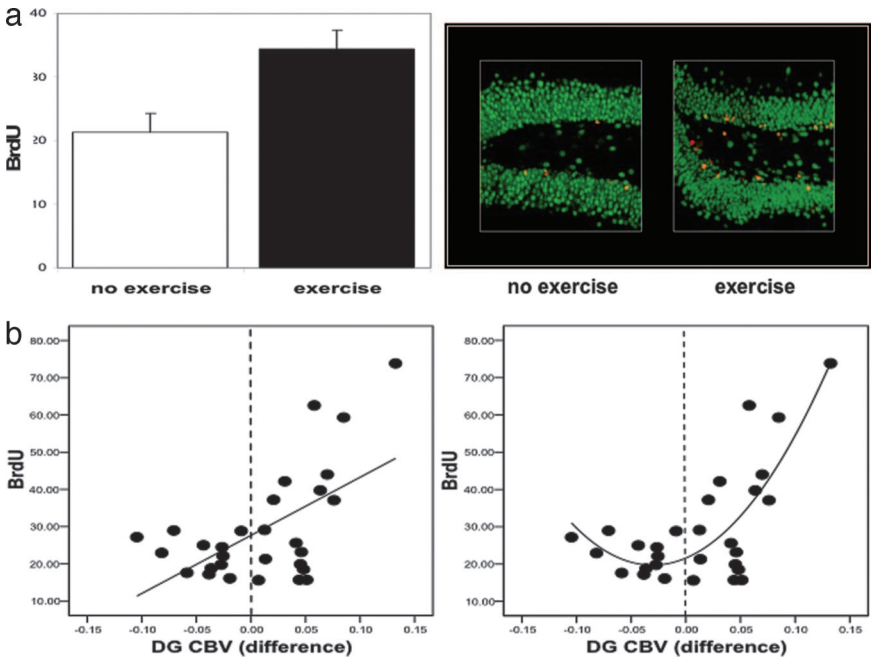


Fig. 2. Exercise-induced increases in dentate gyrus CBV correlate with neurogenesis. (*a*) (*Left*) Exercising mice were found to have more BrdU labeling compared with the no-exercise group. (*Right*) As shown by confocal microscopy, the majority of the new cells were colabeled with NeuN (red, BrdU labeling; green, NeuN; yellow, BrdU/NeuN double labeling). (*b*) (*Left*) A significant linear relationship was found between changes in dentate gyrus CBV and BrdU labeling. (*Right*) A quadratic relationship better fits the data. The vertical stippled line in each plot splits the *x* axis into CBV changes that decreased (left of line) versus those that increased (right of line) with exercise.

memory. Ten subjects were cognitively assessed after exercise, eight of whom were assessed at preexercise baseline.

A repeated-measures ANOVA used to analyze the imaging data showed that the dentate gyrus was the only hippocampal subregion whose CBV significantly increased over time ($F = 12$, $P = 0.006$) (Fig. 3*a*). As in mice, the entorhinal cortex was the only other hippocampal subregion whose CBV increased appreciably over time, although not achieving statistical significance ($F = 4.3$, $P = 0.064$) (Fig. 3*a*). As a group, VO_2max values significantly increased over time ($F = 11.6$, $P = 0.007$) (Fig. 4*a*), and, to confirm that the imaged changes were directly related to exercise and not simply caused by a test–retest effect, we found that individual differences in dentate gyrus CBV were correlated to individual changes in VO_2max ($\beta = 0.662$, $P = 0.027$) (Fig. 4*b*). Importantly, a correlation between CBV and VO_2max was not observed for any other hippocampal subregion, including the entorhinal cortex (Fig. 4*b*), confirming that exercise has a selective effect on dentate gyrus CBV.

Cognitively, individuals performed significantly better on trial 1 learning ($F = 7.0$, $P = 0.027$) after exercise, with a trend toward improvement on all-trial learning ($F = 5.0$, $P = 0.053$) and delayed recall ($F = 5.0$, $P = 0.057$). There was no effect on delayed recognition ($F = 0.19$, $P = 0.67$) or source memory ($F = 0.15$, $P = 0.25$) (Fig. 4*a*). To test that cognitive improvement was related to exercise *per se*, we found that individual changes in trial 1 learning were correlated with individual changes in VO_2max ($\beta = 0.660$, $P = 0.037$). However, because only 8 of the 10 subjects completed preexercise cognitive testing, we repeated the analysis with postexercise cognitive performance scores. Again, we found that changes in VO_2max correlated exclusively with postexercise trial 1 learning ($\beta = 0.70$, $P = 0.026$) (Fig. 4*b*). Additional analyses showed that the correlation between changes in VO_2max and cognition was selective to trial 1 learning (Fig. 4*b*), thereby confirming that, despite apparent increases in other cognitive measures (i.e., delayed recognition, as shown in Fig. 4*a*), this particular measure was selectively influenced by exercise.

Finally, we examined the relationship between cognition and CBV. Among all hippocampal subregions, the correlation between improvements in trial 1 performance and increases in dentate gyrus CBV was the only one that trended toward significance ($\beta = 0.62$, $P = 0.052$). Because of the missing preexercise data, we repeated all of the analyses comparing changes in CBV with postexercise cognition, finding an exclusive correlation between postexercise trial 1 learning and dentate gyrus CBV ($\beta = 0.63$, $P = 0.026$) (Fig. 4*b*).

Discussion

Taken together, our findings show that, within the hippocampal formation, exercise targets the dentate gyrus with regional selectivity. In addition, our results identify that, in mice, dentate gyrus CBV is an imaging correlate of exercise-induced neurogenesis.

Because of its pleiotropic effect on the brain (37, 38), the subregional selectivity that exercise was found to have within the hippocampal formation was an unexpected finding. This observation is particularly interesting in light of studies suggesting that the dentate gyrus is a hippocampal subregion differentially vulnerable to the aging process, as shown in humans (39, 40), nonhuman primates (30, 41), and rodents (30, 42), and that dentate gyrus dysfunction contributes to cognitive aging (30, 40). Interestingly, a growing number of studies have established that exercise ameliorates age-related cognitive decline (43–46). Thus, the effect that exercise was observed to have on the dentate gyrus likely contributes to the reported cognitive benefits exercise has on the aging brain.

Because neurogenesis couples with angiogenesis (7, 8) and angiogenesis, in turn, couples with CBV (18–26), we hypothesized that measures of CBV might provide an *in vivo* correlate of neurogenesis. This hypothesis is based in part on our previous findings that exercise increases both neurogenesis and angiogenesis in young adult rodents (43). We have confirmed this prediction with our high-resolution MRI techniques, which are capable of assessing individual hippocampal subregions, showing that dentate gyrus

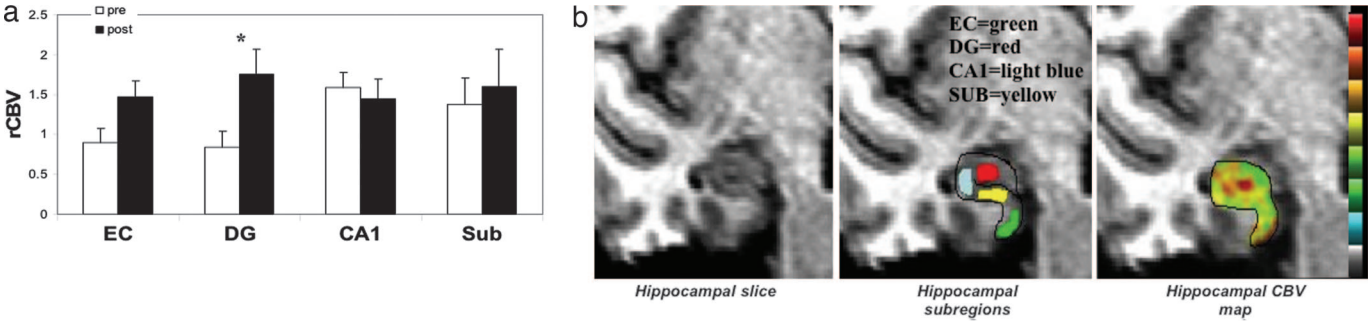


Fig. 3. Exercise selectively increases dentate gyrus CBV in humans. (*a*) Exercise had a selective effect on dentate gyrus CBV. Bar graph shows the mean relative CBV (rCBV) values for each hippocampal subregion before exercise (open bars) and after exercise (filled bars). As in mice, the dentate gyrus was the only hippocampal subregion that showed a significant exercise effect, whereas the entorhinal cortex showed a nonsignificant increase in CBV. (*b*) An individual example. (*Left*) High-resolution MRI slice that visualizes the external morphology and internal architecture of the hippocampal formation. (*Center*) Parcellation of the hippocampal subregions (green, entorhinal cortex; red, dentate gyrus; blue, CA1 subfield; yellow, subiculum). (*Right*) Hippocampal CBV map (warmer colors reflect higher CBV).

CBV selectively correlates with underlying neurogenesis in mice. Blocking the exercise-induced CBV effect with irradiation, a neurogenesis suppressant, further confirms that neurogenesis underlies the observed increases in CBV. However, in the absence of direct measurements, we have not confirmed that angiogenesis is the intermediate factor that accounts for this relationship. Mechanistically, there are many molecules that may be mediating the exercise-induced coupling between neurogenesis and angiogenesis (47), including VEGF (8) or exercise-induced changes in BDNF (48).

The remarkable similarities between the exercise-induced CBV changes in the hippocampal formation of mice and humans suggest that the effect is mediated by similar mechanisms. Of course, in contrast to mice, it is impossible to directly confirm whether the changes in dentate gyrus CBV observed in humans are a reflection of neurogenesis. Nevertheless, previous rodent studies have shown that levels of neurogenesis are coupled with individual differences in degree of exercise (28). Therefore, together with the cross-species similarities on hippocampal CBV, the observation that exercise-induced changes in dentate gyrus CBV correlate with VO_2max supports the interpretation that the human effect is mediated, at least in part, by neurogenesis.

In any case, our results show that *in vivo* imaging can predict levels of neurogenesis in living mice. Furthermore, as demonstrated, the MRI technologies used in these studies are capable of measuring meaningful changes in dentate gyrus CBV over time in both humans and mice. By providing evidence in support of exercise-induced neurogenesis and by introducing the tools needed to achieve this goal, this study shows that it is possible to isolate the specific profile of cognitive changes that are neurogenesis-dependent. Furthermore, the imaging tools presented here are uniquely suited to investigate potential pharmacological modulators of neurogenesis, testing their role in treating depression (49) and in ameliorating the cognitive decline that occurs in all of us as we age (43, 50).

Methods

Exercise. ***Mice.*** A total of 46 C57BL/6 mice (7 weeks old) were used (23 exercising and 23 nonexercising animals). The experimental mice were placed in cages with running wheels (Lafayette Instrument Company, Lafayette, IN). The animals ran voluntarily for 2 weeks. MRI images were acquired at the following time points: week 0 (baseline), week 2 (when exercise was stopped), week 4, and week 6. The thymidine analog BrdU marker was injected i.p. for 7 consecutive days (60 $mg \cdot kg^{-1} \cdot day^{-1}$) during the second week of the experiment. At week 6, the animals were anesthetized and killed in accordance with institutional guidelines.

Human. Eleven subjects (mean age, 33 ranging from 21–45 years; 2 males and 9 females) who fulfilled the American Heart Association criteria for below-average aerobic fitness (VO_2max, <43 for men and <37 for women) were recruited (51). The 11 enrolled subjects engaged in an exercise training protocol for 12 weeks at Columbia University Fitness Center at a frequency of four times a week. Each exercise session lasted ≈1 h: 5 min of low-intensity warm-up on a treadmill or stationary bicycle, 5 min of stretching, 40 min of aerobic training, and 10 min for cool down and stretching. During the 40 min of aerobic activity, subjects were permitted to select from cycling on a stationary ergometer, running on a treadmill, climbing on a StairMaster, or using an elliptical trainer.

VO_2max was measured by a graded exercise test on an Ergoline 800S electronic-braked cycle ergometer (SensorMedics, Anaheim, CA). Each subject began exercising at 30 W for 2 min, and the work rate was continually increased by 30 W each 2 min until VO_2max criteria (respiratory quotient of ≥ 1.1, increases in ventilation without concomitant increases in VO_2, maximum age-predicted heart rate, and/or volitional fatigue) were reached. Minute ventilation was measured by a pneumotachometer connected to a FLO-1 volume transducer module (Physio-Dyne Instrument Corporation, Quogue, NY). Percentages of expired oxygen (O_2) and carbon dioxide (CO_2) were measured by using a paramagnetic O_2 analyzer and an infrared CO_2 analyzer connected to a computerized system (MAX-1; Physio-Dyne Instrument Corporation). These analyzers were calibrated against known medical grade gases. The highest VO_2 value attained during the graded exercise test is considered VO_2max.

In Vivo* Imaging.** ***Mice. Mice were imaged with a 9.4-T Bruker scanner (AVANCBV 400WB spectrometer; Bruker NMR, Billerica, MA) by following a previously described protocol (27). Briefly, axial T2-weighted images were optimally acquired with a fast sequence (time to repeat/effective echo time = 2,000 ms/70 ms; 30-mm i.d. birdcage radio frequency probe; shielded gradient system, 100 g/cm; rapid acquisition with relaxation enhancement factor, 16; field of view, 19.6 mm; acquisition matrix, 256 × 256; no. of slices, 8; slice thickness, 0.6 mm; slice gap, 0.1 mm; number of excitations, 28). Five sets of images were acquired sequentially, each requiring 16 min. The first two sets were precontrast. Gadodiamide was then injected (13 mmol/kg i.p.) through a catheter placed intraperitoneally before imaging. The last three sets corresponded to the postcontrast images. To prevent head motion and reduce anxiety, the animals were anesthetized with isoflurane gas [1.5% (vol/vol) for maintenance at 1 liter per minute of air flow) via a nose cone. Isoflurane was chosen because it induces minimal cerebral hemodynamic change (52). Monitoring of the heart rate, respiratory rate, and oxygen saturation was performed during the whole

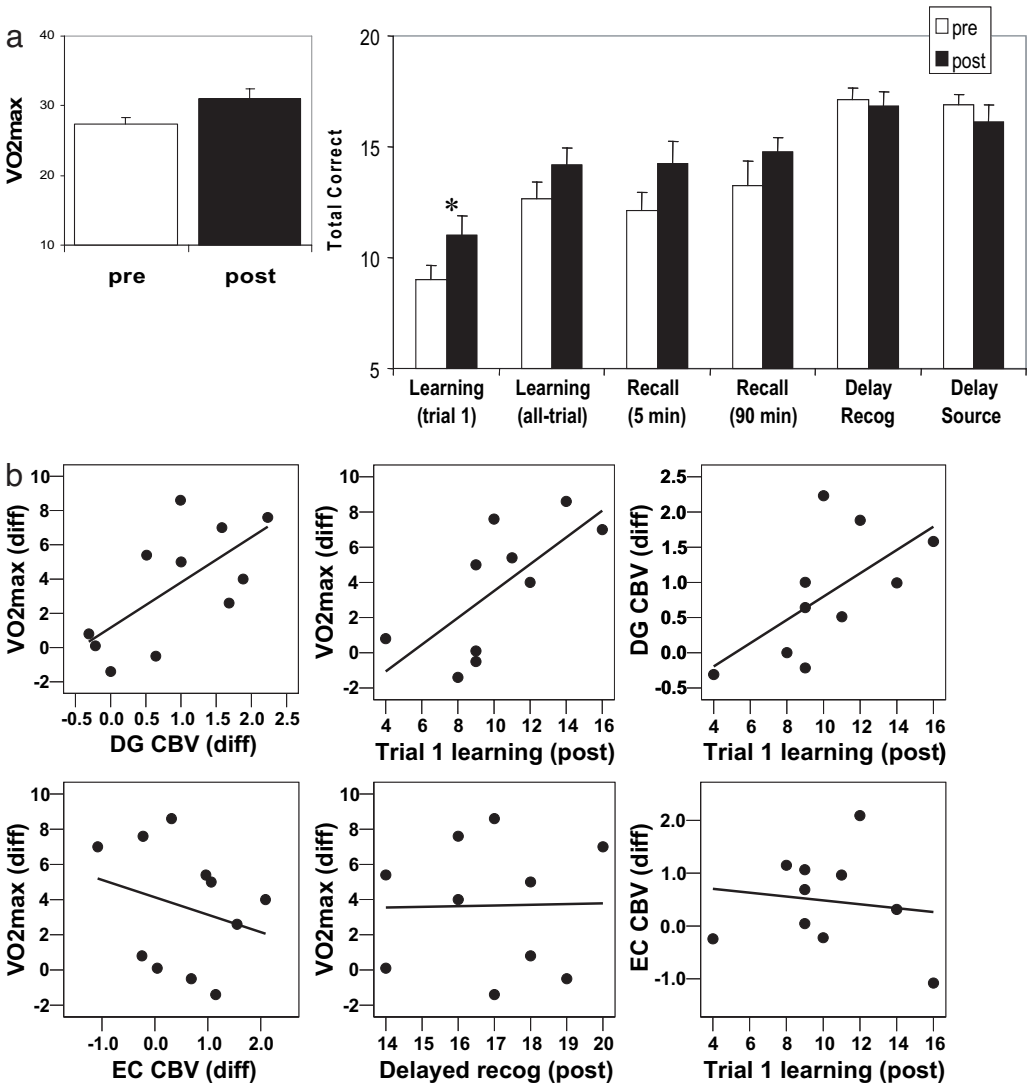


Fig. 4. Exercise-induced increases in dentate gyrus CBV correlate with aerobic fitness and cognition. (*a*) (*Left*) VO_2max, the gold-standard measure of exercise-induced aerobic fitness, increased after exercise. (*Right*) Cognitively, exercise has its most reliable effect on first-trial learning of new declarative memories. (*b*) (*Left*) Exercise-induced changes in VO_2max correlated with changes in dentate gyrus (DG) CBV but not with other hippocampal subregions, including the entorhinal cortex (EC), confirming the selectivity of the exercise-induced effect. (*Center*) Exercise-induced changes in VO_2max correlated with postexercise trial 1 learning but not with other cognitive tasks, including delayed recognition. (*Right*) Postexercise trial 1 learning correlated with exercise-induced changes in dentate gyrus CBV but not with changes in other hippocampal subregions, including the entorhinal cortex.

procedure. Relative CBV was mapped as changes of the transverse relaxation rate (ΔR^2) induced by the contrast agent. When the contrast agent reaches uniform distribution, CBV maps can be measured from steady-state T2-weighted images as CBV $\propto \Delta R^2 = \ln(S_{pre}/S_{post})/TE$, where TE is the effective echo time, S_{pre} is the signal before the contrast administration, and S_{post} is the signal after the contrast agent reaches steady-state. To control for differences in levels of contrast administration, cardiac output, and global blood flow, the derived maps were normalized to the maximum 4-pixel signal value of the posterior cerebral vein. Visualized anatomical landmarks were used together with standard atlases to identify the localization of four hippocampal subregions: the dentate gyrus, the CA3 subfield, the CA1 subfield, and the entorhinal cortex (53). The normalized CBV measurements from each subregion were used for group data analysis.

Human. Subjects were imaged with a 1.5-T scanner Intera scanner (Philips, Amsterdam, The Netherlands). As previously described (30), coronal T1-weighted images (repetition time, 20 ms; echo time, 6 ms; flip angle, 25°; in-plane resolution, 0.86 × 86 mm; slice thickness, 4 mm) were acquired oriented perpendicular to the hippocampal long-axis before and 4 min after i.v. administration of the contrast gadolinium (0.1 mmol/kg). The difference between precontrast and postcontrast images was used to access the regional CBV map. To control for differences in levels of contrast administration, cardiac output, and global blood flow, the derived differences in signal intensity were normalized to the maximum 4-pixel signal value of the sagittal sinus (29). For each subject, the precontrast scan was used to identify the slice with the best visualization of the external morphology and internal architecture of the hippocampal formation. Visualized anatomical landmarks were used together with standard atlases to identify the general locale of four subregions: the dentate gyrus, the CA1 subfield, the subiculum, and the entorhinal cortex (30). The normalized CBV measurements from each subregion were used for group data analysis.

Microscopy. ***Immunohistochemistry.*** Free-floating, 40-μm coronal sections were used in the determination of BrdU labeling. DNA denaturation was conducted by incubation for 1 h at 2 M HCl at

37°C, followed by washing in boric buffer (pH 8.5). After washing, sections were incubated for 30 min in 10% H_2O_2 to eliminate endogenous peroxidases. After blocking with 3% normal donkey serum in 0.2% Triton X-100, sections were incubated with monoclonal anti-BrdU (1:600; Roche, Basel, Switzerland) overnight at 4°C. Sections were then incubated for 1 h at room temperature with the secondary antibody (biotinylated donkey anti-mouse; Jackson ImmunoResearch Laboratories, West Grove, PA) followed by amplification with an avidin–biotin complex (Vector Laboratories, Burlingame, CA), and visualized with diaminobenzidine (Sigma, St. Louis, MO). For double-immunolabeling, free-floating sections were incubated in a mixture of primary antibodies, anti-BrdU (1:600; Roche), and anti-NeuN (1:500; Chemicon, Temecula, CA), raised in different species overnight. For visualization, Alexa Fluor-conjugated appropriate secondary antibodies (1:300; Molecular Probes, Eugene, OR) raised in goats were used for 1 h at room temperature. Blocking serum and primary and secondary antibodies were applied in 0.2% Triton X-100 in PBS. Sections for fluorescent microscopy were mounted on slides in VECTASHIELD (Vector Laboratories). For control of the specificity of immunolabeling, primary antibodies were omitted and substituted with appropriate normal serum. Slides were viewed with a confocal microscope [Radiance 2000 (Bio-Rad, Hercules, CA) mounted on a E800 microscope (Nikon, Tokyo, Japan)]. The images presented are stacks of six to 16 optical sections (step, 1 mm) that were collected individually (in the green and red channels) or simultaneously with precaution against cross-talk between channels. The images were processed with PhotoShop 7.0 (Adobe Systems, San Jose, CA) without contrast and brightness changes in split images.

Quantitation of BrdU labeling. Every sixth section throughout the hippocampus was processed for BrdU immunohistochemistry. Ten sections were used for each animal. All BrdU-labeled cells in the dentate gyrus (granule cell layer and subgranular zone) were counted under a light microscope by an experimenter blinded to the study code. The average number of BrdU-labeled cells was then derived for each animal.

Cognitive testing. Declarative memory was measured with a version of the Rey Auditory Verbal Learning Test (36) modified to increase variability in memory performance among healthy young adults. Twenty nonsemantically or phonemically related words were presented over three learning trials in which the test administrator read the word list and the subject free-recalled as many words as possible. Administration of the three learning trials was immediately followed by one learning trial of a distracter list and then a short delayed free recall of the initial list. After a 90-min delay period, subjects were asked to freely recall words from the initial list and then to freely recall items from the distracter list. After a 24-hr delay period, subjects were contacted by telephone and asked to freely recall items from the initial list and then from the distracter list. They were then administered a forced-choice recognition trial in which they were required to identify the 20 words from the initial learning trial among semantically and phonemically related words, as well as words from the distracter trial. Finally, a source memory trial was administered in which subjects were read a list containing only words from the initial learning list and from the distracter list and were asked to identify from which list each word came. Two forms of the verbal learning test were created, and the administration order was counterbalanced. As in previous studies (54), we measured words correctly recalled on the first trial of the initial learning trials, the average number of words recalled across the three learning trials, the number of words from the initial learning trial that were correctly recalled after a short delay (<5 min), the number of words from the initial learning trial that were correctly recalled after a 90-min delay, the number of items correctly identified on the recognition trial, and the correct number of items identified on the source memory trial.

We thank Kenneth Hess and Fan Hau for help in mouse imaging and Clay Lacefield for helping deliver x-irradiation. This work was supported by grants from the Defense Advanced Research Projects Agency, the National Institutes of Health (AG025161), the New York Heart Association, the Lookout Foundation, and the James S. McDonnell Foundation.

1. Amaral DG (1993) *Curr Opin Neurobiol* 3:225–229.
2. Squire LR, Stark CE, Clark RE (2004) *Annu Rev Neurosci* 27:279–306.
3. Erickson CA, Barnes CA (2003) *Exp Gerontol* 38:61–69.
4. Altman J, Das GD (1965) *J Comp Neurol* 124:319–335.
5. Kaplan MS, Hinds JW (1977) *Science* 197:1092–1094.
6. Eriksson PS, Perfilieva E, Bjork-Eriksson T, Alborn AM, Nordborg C, Peterson DA, Gage FH (1998) *Nat Med* 4:1313–1317.
7. Palmer TD, Willhoite AR, Gage FH (2000) *J Comp Neurol* 425:479–944.
8. Louissaint A, Jr, Rao S, Leventhal C, Goldman SA (2002) *Neuron* 34:945–960.
9. Carmeliet P (2003) *Nat Med* 9:653–660.
10. Risau W (1998) *Kidney Int Suppl* 67:S3–S6.
11. McDonald DM, Choyke PL (2003) *Nat Med* 9:713–725.
12. D'Amore PA, Thompson RW (1987) *Annu Rev Physiol* 49:453–464.
13. Mayr NA, Hawighorst H, Yuh WT, Essig M, Magnotta VA, Knopp MV (1999) *J Magn Reson Imaging* 10:267–276.
14. Taylor JS, Tofts PS, Port R, Evelhoch JL, Knopp M, Reddick WE, Runge VM, Mayr N (1999) *J Magn Reson Imaging* 10:903–907.
15. Bremer C, Mustafa M, Bogdanov A, Jr, Ntziachristos V, Petrovsky A, Weissleder R (2003) *Radiology* 226:214–220.
16. Kerwin W, Hooker A, Spilker M, Vicini P, Ferguson M, Hatsukami T, Yuan C (2003) *Circulation* 107:851–856.
17. van Dijke CF, Brasch RC, Roberts TP, Weidner N, Mathur A, Shames DM, Mann JS, Demsar F, Lang P, Schwickert HC (1996) *Radiology* 198:813–818.
18. Dennie J, Mandeville JB, Boxerman JL, Packard SD, Rosen BR, Weisskoff RM (1998) *Magn Reson Med* 40:793–799.
19. Cha S, Johnson G, Wadghiri YZ, Jin O, Babb J, Zagzag D, Turnbull DH (2003) *Magn Reson Med* 49:848–855.
20. Pathak AP, Rand SD, Schmainda KM (2003) *J Magn Reson Imaging* 18:397–403.
21. Dunn JF, Roche MA, Springett R, Abajian M, Merlis J, Daghlian CP, Lu SY, Makki M (2004) *Magn Reson Med* 51:55–61.
22. Lin TN, Sun SW, Cheung WM, Li F, Chang C (2002) *Stroke* 33:2985–2991.
23. Sugahara T, Korogi Y, Kochi M, Ikushima I, Hirai T, Okuda T, Shigematsu Y, Liang L, Ge Y, Ushio Y, Takahashi M (1998) *Am J Roentgenol* 171:1479–1486.
24. Maia AC, Jr, Malheiros SM, da Rocha AJ, da Silva CJ, Gabbai AA, Ferraz FA, Stavale JN (2005) *Am J Neuroradiol* 26:777–783.
25. Aronen HJ, Pardo FS, Kennedy DN, Belliveau JW, Packard SD, Hsu DW, Hochberg FH, Fischman AJ, Rosen BR (2000) *Clin Cancer Res* 6:2189–2200.
26. Jiang Q, Zhang ZG, Ding GL, Silver B, Zhang L, Meng H, Lu M, Pourabdillah-Nejed DS, Wang L, Savant-Bhonsale S, *et al.* (2006) *NeuroImage* 32:1080–1089.
27. Moreno H, Hua F, Brown T, Small S (2006) *NMR Biomed* 19:535–543.
28. van Praag H, Kempermann G, Gage FH (1999) *Nat Neurosci* 2:266–270.
29. Lin W, Celik A, Paczynski RP (1999) *J Magn Reson Imaging* 9:44–52.
30. Small SA, Chawla MK, Buonocore M, Rapp PR, Barnes CA (2004) *Proc Natl Acad Sci USA* 101:7181–7186.
31. Vissing J, Andersen M, Diemer NH (1996) *J Cereb Blood Flow Metab* 16:729–736.
32. Ide K, Secher NH (2000) *Prog Neurobiol* 61:397–414.
33. van Praag H, Christie BR, Sejnowski TJ, Gage FH (1999) *Proc Natl Acad Sci USA* 96:13427–13431.
34. Mitchell JH, Sproule BJ, Chapman CB (1958) *J Clin Invest* 37:538–547.
35. Wagner PD (1996) *Annu Rev Physiol* 58:21–50.
36. Rey A (1964) *L'Examen Clinique en Psychologie* (Presses Univ de France, Paris).
37. Cotman CW, Berchtold NC (2002) *Trends Neurosci* 25:295–301.
38. Dishman RK, Berthoud HR, Booth FW, Cotman CW, Edgerton VR, Fleshner MR, Gandevia SC, Gomez-Pinilla F, Greenwood BN, Hillman CH, *et al.* (2006) *Obesity (Silver Spring)* 14:345–356.
39. West MJ, Coleman PD, Flood DG, Troncoso JC (1994) *Lancet* 344:769–772.
40. Small SA, Tsai WY, DeLaPaz R, Mayeux R, Stern Y (2002) *Ann Neurol* 51:290–295.
41. Gazzaley AH, Siegel SJ, Kordower JH, Mufson EJ, Morrison JH (1996) *Proc Natl Acad Sci USA* 93:3121–3125.
42. Shetty AK, Turner DA (1999) *Exp Neurol* 158:491–503.
43. van Praag H, Shubert T, Zhao C, Gage FH (2005) *J Neurosci* 25:8680–8685.
44. Kramer AF, Hahn S, Cohen NJ, Banich MT, McAuley E, Harrison CR, Chason J, Vakil E, Bardell L, Boileau RA, Colcombe A (1999) *Nature* 400:418–419.
45. Chan AS, Ho YC, Cheung MC, Albert MS, Chiu HF, Lam LC (2005) *J Am Geriatr Soc* 53:1754–1760.
46. Weuve J, Kang JH, Manson JE, Breteler MM, Ware JH, Grodstein F (2004) *J Am Med Assoc* 292:1454–1461.
47. Emanueli C, Schratzberger P, Kirchmair R, Madeddu P (2003) *Br J Pharmacol* 140:614–619.
48. Neeper SA, Gomez-Pinilla F, Choi J, Cotman CW (1996) *Brain Res* 726:49–56.
49. Santarelli L, Saxe M, Gross C, Surget A, Battaglia F, Dulawa S, Weisstaub N, Lee J, Duman R, Arancio O, *et al.* (2003) *Science* 301:805–809.
50. Small SA, Stern Y, Tang M, Mayeux R (1999) *Neurology* 52:1392–1396.
51. Fletcher GF, Balady G, Froelicher VF, Hartley LH, Haskell WL, Pollock ML (1995) *Circulation* 91:580–615.
52. Lei H, Grinberg O, Nwaigwe CI, Hou HG, Williams H, Swartz HM, Dunn JF (2001) *Brain Res* 913:174–179.
53. Paxinos G, Franklin K (2001) *The Mouse Brain in Stereotaxic Coordinates* (Academic, Philadelphia).
54. Van der Elst W, van Boxtel MP, van Breukelen GJ, Jolles J (2005) *J Int Neuropsychol Soc* 11:290–302.
55. Santarelli L, Saxe M, Gross C, Surget A, Battaglia F, Dulawa S, Weisstaub N, Lee J, Duman R, Arancio O, *et al.* (2003) *Science* 301:805–809.

Differential Diagnosis of Cushing's Syndrome

ABSTRACT

The differential diagnosis of Cushing's syndrome requires careful multidisciplinary interaction with a number of specialities, co-ordinated through endocrine centres with good experience of this condition. It is essential that the diagnosis of Cushing's syndrome be fully established before differential diagnosis is attempted. The endocrinologist needs to be aware of the pitfalls and advantages of the tests in use. We discuss the approach to the differential diagnosis of this challenging condition. **(Arq Bras Endocrinol Metab 2007;51/8:1199-1206)**

Keywords: Cushing's syndrome; Cortisol; ACTH; Dexamethasone; Adrenal; Pituitary

RESUMO

Diagnóstico Diferencial da Síndrome de Cushing.
O diagnóstico diferencial da síndrome de Cushing requer uma interação multidisciplinar cuidadosa entre várias especialidades, coordenadas através de centros de endocrinologia com boa experiência nessa condição. É essencial que o diagnóstico da síndrome de Cushing seja estabelecido antes da tentativa de diagnóstico diferencial. O endocrinologista precisa estar atento às possíveis falhas e vantagens dos testes empregados. Nós discutiremos a abordagem do diagnóstico diferencial nessa condição desafiadora. **(Arq Bras Endocrinol Metab 2007;51/8:1199-1206)**

Descritores: Síndrome de Cushing; Cortisol; ACTH; Dexametasona; Adrenal; Hipófise

revisão

JOHN NEWELL-PRICE
ASHLEY B. GROSSMAN

Academic Unit of Diabetes, Endocrinology & Metabolism, School of Medicine and Biomedical Science, The University of Sheffield, Sheffield, UK (JN-P), and Dept. of Endocrinology, William Harvey Research Institute, Barts and the London, Queen Mary's School of Medicine and Dentistry, London EC1A 7BE, UK (ABG).

ONCE THE DIAGNOSIS of the hypercortisolism of Cushing's syndrome is firmly established, attention can be turned to establishing the cause. This is an important point as all the tests used in the differential diagnosis depend on the prior demonstration of hypercortisolism. If there is any doubt diagnostic tests should be repeated, if needed on several occasions, before continuing on the differential diagnostic pathways. We give an overview of the differential diagnostic pathway, of which the individual parts are also discussed in detail in other articles in this edition.

CAUSES OF CUSHING'S SYNDROME

To understand the limitations of the various tests used in the differential diagnosis of Cushing's syndrome it is important to know the approximate likelihood of the cause. Endogenous Cushing's syndrome is com-

Recebido em 30/07/07
Aceito em 10/08/07

moner in women than men and is divided into adrenocorticotrophin (ACTH)-dependent, and ACTH-independent, causes (table 1). Overall, ACTH-dependent causes account for approximately 80–85% of cases, and of these 80% are due to pituitary adenomas (*Cushing's disease*), with the remaining 20% or so due to the ectopic ACTH syndrome (1,2). Ectopic ACTH secretion is most commonly from small cell carcinoma of the lung and bronchial carcinoid tumours, but may also occur with almost any endocrine tumour from many different organs (e.g. phaeochromocytoma, pancreatic neuroendocrine tumours, gut carcinoids) (table 2) (3,5). When the ectopic ACTH syndrome is due to small cell carcinoma of the lung it may have a rapid onset with severe features of muscle weakness, but without the classical Cushing's phenotype. In contrast, the clinical phenotype (and some biochemical features) of carcinoid tumours may be indistinguishable from that of Cushing's disease (3-6). In ectopic ACTH syndrome when the source of ACTH is apparent on simple imaging, this is referred to as the 'overt' ectopic ACTH syndrome and may be due to rapidly progressive small cell lung cancer or a more indolent source. If the source of ectopic ACTH is found only after extensive imaging this is referred to as 'covert' ectopic ACTH, whilst if the source is not found despite extensive imaging, this is termed 'occult'.

ACTH-independent Cushing's syndrome is due in most cases to a unilateral tumour: adrenal adenoma in 60% and carcinoma in 40% of cases. Very rare adrenal causes of Cushing's syndrome are ACTH-independent macronodular adrenal hyperplasia (AIMAH), primary pigmented nodular adrenal disease (PPNAD) (either as isolated disease or as part of Carney complex), and the McCune Albright syndrome (2,7,8).

ESTABLISHING THE AETIOLOGY OF CUSHING'S SYNDROME (figure 1)

Measurement of plasma ACTH

The differential diagnosis of Cushing's syndrome is best done in major referral centres. Investigation will vary depending upon the availability of the biochemical tests and imaging detailed below. The first step is to measure plasma ACTH. Levels consistently below 5 pg/ml indicate ACTH-independent Cushing's syndrome. In this case the next step is imaging the adrenal glands with CT or MRI. Levels of ACTH persistently above 15 pg/ml almost always reflect ACTH-dependent pathologies and require investigation, as detailed below. Plasma ACTH values between 5–15 pg/ml need careful interpretation since some patients with Cushing's disease and adrenal pathologies may have intermediate values (2,9,10). It is essential that the sample for plasma ACTH is separated rapidly and stored at -40°C to avoid degradation and a falsely low result. A positive CRH test (see below) can be useful to demonstrate ACTH-dependent hypercortisolism in an occasional patient with Cushing's disease with low baseline ACTH plasma levels.

ACTH-independent Cushing's syndrome

When adrenal adenoma, carcinoma or AIMAH is the cause of ACTH-independent Cushing's syndrome, the anatomical cause is invariably visible on imaging with CT (11). In cases of unilateral adrenal adenomas and carcinomas causing Cushing's syndrome it is usual for the contra-lateral adrenal gland to undergo atrophy in the prolonged absence of plasma ACTH. One simple means of assessing the size of the limbs of the adrenal gland is to compare them to the adjacent crus of the diaphragm on CT — if larger than the width of the

Table 1. Aetiology of Cushing's syndrome.

Cause of Cushing's syndrome	%	F:M
ACTH-dependent		
Cushing's disease	70%	3.5:1
Ectopic ACTH syndrome	10%	1:1
Unknown source of ACTH*	5%	5:1
ACTH-independent		
Adrenal adenoma	10%	4:1
Adrenal carcinoma	5%	1:1
Macronodular hyperplasia (AIMAH)	< 2%	
Primary pigmented nodular adrenal disease (PPNAD)	< 2%	
McCune Albright syndrome	< 2%	

* Patients may ultimately prove to have Cushing's disease.

Table 2. Aetiology of ectopic ACTH secretion.

	Institution		
	NIH	**Barts**	**Sao Paulo**
Bronchial carcinoid tumour	35	12	10
Small cell lung carcinoma	3	7	
Pulmonary Tumourlets		1	
Thymic carcinoid tumour	5	2	4
Medullary thyroid carcinoma	2	3	
Pancreatic carcinoid tumour	1	3	3
Gastrinoma	6	3	
Appendix	1		
Other GI Neuroendocrine tumours (NET)		13	
Mesothelioma	1		
Colonic carcinoma		2	
Phaeochromocytoma	5	1	5
Olfactory Ethesioneuroblastoma		1	
Disseminated carcinoid tumour		2	
Lymph Node NET		2	
Glomus tumour			1
Unknown	17	2	2
Overt	46	26	20
Covert	23	9	3
Occult	17	5	2
Total	**90**	**40**	**25**

Data derived from: NIH – Ilias et al. (ref. 5); Barts – Isidori et al. (ref. 4); Sao Paulo – Salgado et al. (ref. 3).

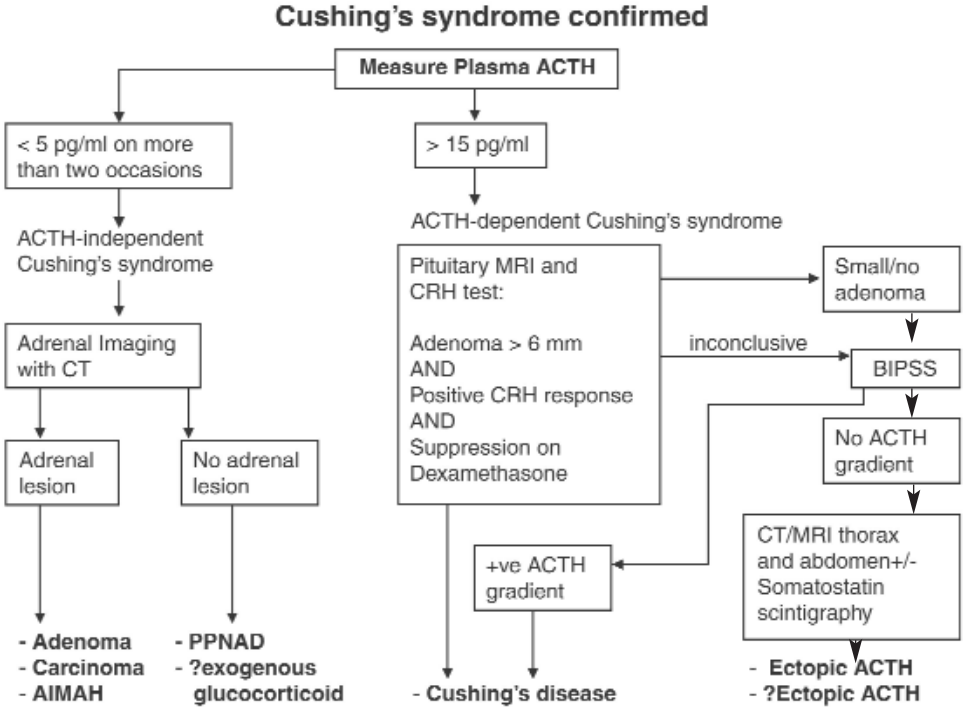


Figure 1. Diagnosis of cause of Cushing's syndrome.
AIMAH: ACTH-independent macronodular hyperplasia, PPNAD: primary pigmented nodular adrenal disease, BIPSS: bilateral inferior petrosal sinus sampling, SCLC: small cell lung cancer.

diaphragm, as seen on CT, hyperplasia is possible and caution is needed to ensure that bilateral disease is not inadvertently characterised as unilateral.

In PPNAD, the adrenal glands may appear normal. Thus, in an established diagnosis of ACTH-independent Cushing's syndrome with normal appearances of the adrenal glands on imaging assessment of other features of Carney complex (lentigines, myxoma), and genetic testing for mutations of *PRKAR1A* may be of benefit as a diagnostic procedure. Exogenous glucocorticoid ingestion (especially hydrocortisone but also synthetic glucocorticoids that cross react in the cortisol assay) should also be reconsidered in this setting.

ACTH-dependent Cushing's syndrome

Overview

Differentiating between pituitary and non-pituitary sites of excess ACTH secretion may be a considerable challenge in clinical endocrinology. Neuroendocrine (carcinoid) tumours may be clinically indistinguishable from Cushing's disease. Moreover, as they are usually small they are frequently difficult to identify with imaging, especially if radiological (pituitary, thoracic, pancreatic) "incidentalomas" complicate interpretation. It is for these reasons that biochemical evaluation is relied upon rather than imaging, to differentiate between pituitary and non-pituitary causes (1-5). It is important to note that in women with ACTH-dependent Cushing's syndrome 9 out of 10 cases will be due to Cushing's disease. It is against this pre-test likelihood that the performance of any test needs to be judged. If a patient has ACTH-dependent Cushing's syndrome, with responses *both* on dexamethasone suppression *and* CRH testing suggesting pituitary disease, and the pituitary MRI scan shows an isolated lesion of 6 mm or more, most will regard the diagnosis of Cushing's disease to have been made. A major problem is that up to 40% of patients with proven Cushing's disease have normal pituitary MRI scans (9). The results of CRH and dexamethasone tests, and pituitary MRI, should be considered together, and bilateral inferior petrosal sinus sampling (BIPSS) is usually recommended unless there is a clear diagnosis (figure 2).

Plasma potassium

High levels of cortisol may either saturate the 11β-hydroxysteroid dehydrogenase type II enzyme in the kidney, or decrease expression of this enzyme, allowing cortisol to act even more as a mineralocorticoid (12). The commonest cause of hypokalaemia is the ectopic ACTH syndrome, but it is also present in those patients with Cushing's disease with extremely high cortisol production (approximately 10% of patients) (2).

Plasma ACTH

The circulating levels of plasma ACTH are usually lower in patients with Cushing's disease compared to those with ectopic secretion. There is, however, considerable overlap and this means that the absolute levels cannot be used reliably to discriminate between groups.

DYNAMIC NON-INVASIVE TESTS

High-dose dexamethasone-suppression test (HDDST)

The high dose dexamethasone suppression tests have been in widespread use for many years. In most centres either dexamethasone 2 mg given every 6 hours for 48 hours, or a single 8 mg dose is given with serum or urinary cortisol measured at the beginning and end of the test. Measurement of serum cortisol is easier and superior to measurement of urinary free cortisol or steroid metabolites. The HDDST relies upon the relative sensitivity of pituitary corticotroph adenomas to the effects of glucocorticoids, compared to the resistance exhibited by non-pituitary tumours. Approximately 80% of patients with Cushing's disease will demonstrate suppression of the serum cortisol to a value of < 50% of the basal level (2). This is less than the pre-test likelihood of Cushing's disease, and thus, by itself, the high dose dexamethasone suppression test has little diagnostic utility (13). Moreover, in very large series we have shown that when utilising the 48-hour low-dose dexamethasone-suppression test, if there has already been the demonstration of suppression of serum cortisol by more than 30%, there is no further advantage to going on to a high dose dexamethasone suppression test (14). Therefore, overall, we would not recommend the continued routine use of the HDDST except when BIPSS is not available.

The corticotrophin-releasing hormone (CRH) test

The basis of the corticotropin-releasing hormone test is that CRH stimulates corticotroph tumour cells in the pituitary to release ACTH and hence increase serum cortisol concentrations, whilst responses are uncommon in the ectopic ACTH syndrome. This is because of the frequent expression of the CRH-1 receptor in corticotroph tumours, which is rare in

ectopic tumours. Recombinant human or ovine-sequence CRH is administered as an intravenous bolus dose of either 1 μg/kg, or in adults 100 μg IV. The ovine-sequence CRH test has a sensitivity of 93% for Cushing's disease based on ACTH responses at 15 and 30 minutes (15). Using more extensive sampling (up to 90 minutes) and a more stringent cut-off point of a 50% increment in plasma ACTH, ovine-sequence CRH had a sensitivity of 86% for Cushing's disease (16). An almost identical sensitivity is found for the human sequence peptide sampling at the same time points (17). Responses to both CRH testing and the high-dose dexamethasone suppression test are also more frequently discordant in patients with Cushing's disease secondary to a pituitary macroadenoma (18).

Desmopressin testing

Since the V3 receptor is expressed in pituitary and many ectopic tumours secreting ACTH (19-22), the desmopressin test is of limited utility in the differential diagnosis of ACTH-dependent Cushing's syndrome (23,24). Similarly, a combined test using CRH and desmopressin has been used (25), but larger series have suggested that there remains overlap between responses in patients with Cushing's disease and the ectopic ACTH syndrome (26) due to the more frequent expression of the V3 receptor in ectopic tumours.

INVASIVE TESTING

Bilateral inferior petrosal sinus sampling (BIPSS)

Sampling of the gradient of ACTH from the pituitary to the periphery is the most reliable means for discriminating between pituitary and non-pituitary sources of ACTH. When there is a central source of ACTH there is a gradient of the value of plasma ACTH compared to the simultaneous peripheral sample, since the pituitary effluent drains via the cavernous sinuses to the petrosal sinuses and then the jugular bulb. BIPSS is a highly skilled and invasive technique, requiring placement of catheters in both inferior petrosal sinuses. Catheter position and venous anatomy require confirmation by venography, as non-uniform drainage is not uncommon. The diagnostic accuracy of the test requires the administration of CRH. A basal central: peripheral ratio of > 2:1 or a CRH stimulated ratio of > 3:1 is consistent with Cushing's disease (27). The combined data for many series indicates a sensitivity and a specificity of 94% (28). Small series data have suggested that these false negative responses can be identified by simultaneous sampling of prolactin to correct values in ACTH (29,30). It is possible that falsely positive results may be caused by inadequate suppression of the normal corticotrophs and the duration and amount of hypercortisolism should be assessed prior to the test. Where CRH is unobtainable or too costly, desmopressin offers a reasonable alternative, but few patients with ectopic ACTH secretion have been studied in this way (31,32). Using both peptides during BIPSS in combination has also been recently reported and appears effective, but again few patients with ectopic ACTH have been studied (33).

In adults, BIPSS is only 70% accurate for lateralisation of the source of ACTH within the pituitary gland (2,10), but in children it may have greater accuracy for this purpose than MRI (34). Sampling from the cavernous sinuses directly does not improve accuracy (35).

Internal jugular venous sampling

Sampling from the internal jugular vein has been proposed as a simplified procedure. Direct comparison in the same patients has shown this to be inferior to BIPSS (36). One application of this test, however, may be in centres of limited sampling experience, and BIPSS (performed in a referral centre) reserved if the results are negative (37).

IMAGING

Adrenal

CT gives the best resolution of adrenal anatomy. It is important to note that in ACTH-dependent Cushing's syndrome nodules may occur and adrenal hyperplasia is not always symmetrical. This can cause diagnostic confusion with a unilateral primary adrenal cause if the biochemistry is not strictly assessed. Conversely, in 30% of Cushing's disease the adrenal glands appear normal, whilst in ectopic ACTH the adrenals are virtually always homogeneously enlarged (38).

Pituitary

Up to 40% of corticotroph adenomas causing Cushing's disease in adults are not visible on standard sequence 1.5 tesla MRI scanning (9). This fact is further complicated by a 10% rate of pituitary incidentalomas in the normal population (39), emphasising the need for careful biochemical discrimination of pituitary from non-pituitary sources of ACTH. Therefore, in the absence of a pituitary macroadenoma, an abnormal MRI is not conclusive evidence in favour of Cushing's disease.

Corticotroph tumours that are visible usually fail to enhance following gadolinium on T1-weighted imaging. The use of dynamic MRI, with the administration of intravenous contrast media and rapid sequence acquisition following this, does not improve the overall diagnostic rate. However, spoiled gradient sequences may have greater sensitivity in adults (40) and children (41).

Imaging in the ectopic ACTH syndrome

If clinical and biochemical assessments suggest a non-pituitary source of ACTH, then detailed further imaging studies are needed. Axial imaging with thin cut multi-slice CT of thorax and abdomen, and/or MRI of the thorax, has the highest detection rate for the ectopic ACTH syndrome. Small neuroendocrine tumours frequently express somatostatin receptors and may be disclosed on somatostatin receptor scintigraphy. However, while standard somatostatin scintigraphy may confirm functionality for a lesion found on axial imaging, it has only rarely been shown to disclose truly "occult" tumours that are not visible on CT (2,4,5,42-44). Using higher than standard doses of radionuclide may, in some cases, disclose lesions that were otherwise negative on imaging. In patients with recurrent disease, somatostatin scintigraphy may be useful for follow-up (45) as it has a low false-positive rate (46). Positron emission tomography (PET) with 18-flurodeoxyglucose (FDG) is of little benefit as such tumours are usually of low metabolic activity (47). The use of ^{11}C-5-hydroxytryptophan with PET has been proposed as a universal imaging technique for neuroendocrine tumours, but few patients have been studied (48) and further experience is required to determine its utility. There is also increasing interest in ^{68}Ga-octreotate PET scanning. Despite extensive investigation, the cause of ACTH may remain 'occult' in 5–15% of patients, and these require continued follow-up. Over time the number remaining undiagnosed reduces as tests are repeated, although establishing the cause occasionally may take up to 20 years of follow up (49).

CONCLUSIONS

The differential diagnosis of Cushing's syndrome requires multidisciplinary interaction with a number of specialities, co-ordinated through endocrine centres with good experience of this condition. It is essential the diagnosis of Cushing's syndrome be fully established before differential diagnosis is attempted. The endocrinologist needs to be aware of the pitfalls of the tests in use, and that these may need repeating over time if the source of ACTH is not found at initial diagnostic workup.

REFERENCES

1. Newell-Price J, Bertagna X, Grossman AB, Nieman LK. Cushing's syndrome. **Lancet 2006**;367:1605-17.
2. Newell-Price J, Trainer P, Besser M, Grossman A. The diagnosis and differential diagnosis of Cushing's syndrome and pseudo-Cushing's states. **Endocr Rev 1998**;19:647-72.
3. Salgado LR, Fragoso MC, Knoepfelmacher M, Machado MC, Domenice S, Pereira MA, et al. Ectopic ACTH syndrome: our experience with 25 cases. **Eur J Endocrinol 2006**;155:725-33.
4. Isidori AM, Kaltsas GA, Pozza C, Frajese V, Newell-Price J, Reznek RH, et al. The ectopic adrenocorticotropin syndrome: clinical features, diagnosis, management, and long-term follow-up. **J Clin Endocrinol Metab 2006**;91:371-7.
5. Ilias I, Torpy DJ, Pacak K, Mullen N, Wesley RA, Nieman LK. Cushing's syndrome due to ectopic corticotropin secretion: twenty years' experience at the National Institutes of Health. **J Clin Endocrinol Metab 2005**;90:4955-62.
6. Aniszewski JP, Young WF, Jr., Thompson GB, Grant CS, van Heerden JA. Cushing syndrome due to ectopic adrenocorticotropic hormone secretion. **World J Surg 2001**;25:934-40.
7. Nieman LK. **Cushing's syndrome**. Philadephia: W.B. Saunders Company, **2001**.
8. Lacroix A, Ndiaye N, Tremblay J, Hamet P. Ectopic and abnormal hormone receptors in adrenal Cushing's syndrome. **Endocr Rev 2001**;22:75-110.
9. Invitti C, Pecori Giraldi F, de Martin M, Cavagnini F. Diagnosis and management of Cushing's syndrome: results of an Italian multicentre study. Study Group of the Italian Society of Endocrinology on the Pathophysiology of the Hypothalamic-Pituitary-Adrenal Axis. **J Clin Endocrinol Metab 1999**;84:440-8.
10. Arnaldi G, Angeli A, Atkinson AB, Bertagna X, Cavagnini F, Chrousos GP, et al. Diagnosis and complications of Cushing's syndrome: a consensus statement. **J Clin Endocrinol Metab 2003**;88:5593-602.
11. Rockall AG, Babar SA, Sohaib SA, Isidori AM, Diaz-Cano S, Monson JP, et al. CT and MR imaging of the adrenal glands in ACTH-independent Cushing syndrome. **Radiographics 2004**;24:435-52.
12. Stewart PM, Walker BR, Holder G, O'Halloran D, Shackleton CH. 11 beta-Hydroxysteroid dehydrogenase activity in Cushing's syndrome: explaining the mineralocorticoid excess state of the ectopic adrenocorticotropin syndrome. **J Clin Endocrinol Metab 1995**;80(12):3617-20.
13. Aron DC, Raff H, Findling JW. Effectiveness versus efficacy: the limited value in clinical practice of high dose dexamethasone suppression testing in the differential diagnosis of adrenocorticotropin-dependent Cushing's syndrome. **J Clin Endocrinol Metab 1997**;82(6):1780-5.
14. Isidori AM, Kaltsas GA, Mohammed S, Morris DG, Jenkins P, Chew SL, et al. Discriminatory value of the low-dose dexamethasone suppression test in establishing the diagnosis and differential diagnosis of Cushing's syndrome. **J Clin Endocrinol Metab 2003**;88(11):5299-306.
15. Nieman LK, Oldfield EH, Wesley R, Chrousos GP, Loriaux DL, Cutler GB, Jr. A simplified morning ovine corticotropin-releasing hormone stimulation test for the differential diagnosis of adrenocorticotropin-dependent Cushing's syndrome. **J Clin Endocrinol Metab 1993**;77(5):1308-12.

16. Reimondo G, Paccotti P, Minetto M, Termine A, Stura G, Bergui M, et al. The corticotrophin-releasing hormone test is the most reliable noninvasive method to differentiate pituitary from ectopic ACTH secretion in Cushing's syndrome. **Clin Endocrinol (Oxf) 2003**;58(6):718-24.
17. Newell-Price J, Morris DG, Drake WM, Korbonits M, Monson JP, Besser GM, et al. Optimal response criteria for the human CRH test in the differential diagnosis of ACTH-dependent Cushing's syndrome. **J Clin Endocrinol Metab 2002**; 87(4):1640-5.
18. Woo YS, Isidori AM, Wat WZ, Kaltsas GA, Afshar F, Sabin I, et al. Clinical and biochemical characteristics of adrenocorticotropin-secreting macroadenomas. **J Clin Endocrinol Metab 2005**;90(8):4963-9.
19. Arlt W, Dahia PL, Callies F, Nordmeyer JP, Allolio B, Grossman AB, et al. Ectopic ACTH production by a bronchial carcinoid tumour responsive to desmopressin in vivo and in vitro. **Clin Endocrinol (Oxf) 1997**;47(5):623-7.
20. de Keyzer Y, Rene P, Lenne F, Auzan C, Clauser E, Bertagna X. V3 vasopressin receptor and corticotropic phenotype in pituitary and nonpituitary tumors. **Horm Res 1997**;47(4-6): 259-62.
21. Dahia PL, Ahmed-Shuaib A, Jacobs RA, Chew SL, Honegger J, Fahlbusch R, et al. Vasopressin receptor expression and mutation analysis in corticotropin-secreting tumors. **J Clin Endocrinol Metab 1996**;81(5):1768-71.
22. Dahia PL, Grossman AB. The molecular pathogenesis of corticotroph tumors. **Endocr Rev 1999**;20(2):136-55.
23. Newell-Price J. The desmopressin test and Cushing's syndrome: current state of play. **Clin Endocrinol (Oxf) 1997**;47(2):173-4.
24. Tsagarakis S, Tsigos C, Vasiliou V, Tsiotra P, Kaskarelis J, Sotiropoulou C, et al. The desmopressin and combined CRH-desmopressin tests in the differential diagnosis of ACTH-dependent Cushing's syndrome: constraints imposed by the expression of V2 vasopressin receptors in tumors with ectopic ACTH secretion. **J Clin Endocrinol Metab 2002**; 87(4):1646-53.
25. Newell-Price J, Perry L, Medbak S, Monson J, Savage M, Besser M, et al. A combined test using desmopressin and corticotropin-releasing hormone in the differential diagnosis of Cushing's syndrome. **J Clin Endocrinol Metab 1997**; 82(1):176-81.
26. Tsagarakis S, Kaskarelis IS, Kokkoris P, Malagari C, Thalassinos N. The application of a combined stimulation with CRH and desmopressin during bilateral inferior petrosal sinus sampling in patients with Cushing's syndrome. **Clin Endocrinol (Oxf) 2000**;52(3):355-61.
27. Oldfield EH, Doppman JL, Nieman LK, Chrousos GP, Miller DL, Katz DA, et al. Petrosal sinus sampling with and without corticotropin-releasing hormone for the differential diagnosis of Cushing's syndrome. **N Engl J Med 1991**; 325(13): 897-905.
28. Lindsay JR, Nieman LK. Differential diagnosis and imaging in Cushing's syndrome. **Endocrinol Metab Clin North Am 2005**;34(2):403-21.
29. McNally PG, Bolia A, Absalom SR, Falconer-Smith J, Howlett TA. Preliminary observations using endocrine markers of pituitary venous dilution during bilateral simultaneous inferior petrosal sinus catheterization in Cushing's syndrome: is combined CRF and TRH stimulation of value? **Clin Endocrinol (Oxf) 1993**;39(6):681-6.
30. Findling JW, Kehoe ME, Raff H. Identification of patients with Cushing's disease with negative pituitary adrenocorticotropin gradients during inferior petrosal sinus sampling: prolactin as an index of pituitary venous effluent. **J Clin Endocrinol Metab 2004**;89(12):6005-9.
31. Machado MC, de Sa SV, Domenice S, Fragoso MC, Puglia P Jr, Pereira MA, et al. The role of desmopressin in bilateral and simultaneous inferior petrosal sinus sampling for differential diagnosis of ACTH-dependent Cushing's syndrome. **Clin Endocrinol (Oxf) 2007**;66(1):136-42.
32. Gannage-Yared MH, Slaba S, Rizk T, Chidiac Wehbe RM. Use of desmopressin as an alternative to corticotropinreleasing hormone during inferior petrosal sinus sampling in a child with Cushing's disease. **J Endocrinol Invest 2007**; 30(5):434-8.
33. Tsagarakis S, Vassiliadi D, Kaskarelis IS, Komninos J, Souvatzoglou E, Thalassinos N. The application of the combined corticotropin-releasing hormone plus desmopressin stimulation during petrosal sinus sampling is both sensitive and specific in differentiating patients with Cushing's disease from patients with the occult ectopic adrenocorticotropin syndrome. **J Clin Endocrinol Metab 2007**;92(6):2080-6.
34. Lienhardt A, Grossman AB, Dacie JE, Evanson J, Huebner A, Afshar F, et al. Relative contributions of inferior petrosal sinus sampling and pituitary imaging in the investigation of children and adolescents with ACTH-dependent Cushing's syndrome. **J Clin Endocrinol Metab 2001**; 86(12):5711-4.
35. Liu C, Lo JC, Dowd CF, Wilson CB, Kunwar S, Aron DC, et al. Cavernous and inferior petrosal sinus sampling in the evaluation of ACTH-dependent Cushing's syndrome. **Clin Endocrinol (Oxf) 2004**;61(4):478-86.
36. Erickson D, Huston J 3rd, Young WF Jr, Carpenter PC, Wermers RA, Bonelli FS, et al. Internal jugular vein sampling in adrenocorticotropic hormone-dependent Cushing's syndrome: a comparison with inferior petrosal sinus sampling. **Clin Endocrinol (Oxf) 2004**;60(4):413-9.
37. Ilias I, Chang R, Pacak K, Oldfield EH, Wesley R, Doppman J, et al. Jugular venous sampling: an alternative to petrosal sinus sampling for the diagnostic evaluation of adrenocorticotropic hormone-dependent Cushing's syndrome. **J Clin Endocrinol Metab 2004**;89(8):3795-800.
38. Sohaib SA, Hanson JA, Newell-Price JD, Trainer PJ, Monson JP, Grossman AB, et al. CT appearance of the adrenal glands in adrenocorticotrophic hormone-dependent Cushing's syndrome. **AJR Am J Roentgenol 1999**;172(4):997-1002.
39. Hall WA, Luciano MG, Doppman JL, Patronas NJ, Oldfield EH. Pituitary magnetic resonance imaging in normal human volunteers: occult adenomas in the general population. **Ann Intern Med 1994**;120(10):817-20.
40. Patronas N, Bulakbasi N, Stratakis CA, Lafferty A, Oldfield EH, Doppman J, et al. Spoiled gradient recalled acquisition in the steady state technique is superior to conventional postcontrast spin echo technique for magnetic resonance imaging detection of adrenocorticotropin-secreting pituitary tumors. **J Clin Endocrinol Metab 2003**; 88(4):1565-9.
41. Batista D, Courkoutsakis NA, Oldfield EH, Griffin KJ, Keil M, Patronas NJ, et al. Detection of adrenocorticotropin-secreting pituitary adenomas by magnetic resonance imaging in children and adolescents with Cushing disease. **J Clin Endocrinol Metab 2005**;90(9):5134-40.
42. Lamberts SW, de Herder WW, Krenning EP, Reubi JC. A role of (labeled) somatostatin analogs in the differential diagnosis and treatment of Cushing's syndrome. **J Clin Endocrinol Metab 1994**;78(1):17-9.
43. Wajchenberg BL, Mendonça BB, Liberman B, Pereira MA, Carneiro PC, Wakamatsu A, et al. Ectopic adrenocorticotropic hormone syndrome. **Endocr Rev 1994**;15(6):752-87.
44. Phlipponneau M, Nocaudie M, Epelbaum J, De Keyzer Y, Lalau JD, Marchandise X, et al. Somatostatin analogs for the localization and preoperative treatment of an adrenocorticotropin-secreting bronchial carcinoid tumor. **J Clin Endocrinol Metab 1994**;78(1):20-4.
45. Granberg D, Sundin A, Janson ET, Oberg K, Skogseid B, Westlin JE. Octreoscan in patients with bronchial carcinoid tumours. **Clin Endocrinol (Oxf) 2003**;59(6):793-9.
46. Tsagarakis S, Christoforaki M, Giannopoulou H, Rondogianni F, Housianakou I, Malagari C, et al. A reappraisal of the utility of somatostatin receptor scintigraphy in patients with ectopic adrenocorticotropin Cushing's syndrome. **J Clin Endocrinol Metab 2003**;88(10):4754-8.

47. Pacak K, Ilias I, Chen CC, Carrasquillo JA, Whatley M, Nieman LK. The role of [(18)F]fluorodeoxyglucose positron emission tomography and [(111)In]-diethylenetriaminepentaacetate-D-Phe-pentetreotide scintigraphy in the localization of ectopic adrenocorticotropin-secreting tumors causing Cushing's syndrome. **J Clin Endocrinol Metab 2004**;89(5):2214-21.
48. Orlefors H, Sundin A, Garske U, Juhlin C, Oberg K, Skogseid B, et al. Whole-body (11)C-5-hydroxytryptophan positron emission tomography as a universal imaging technique for neuroendocrine tumors: comparison with somatostatin receptor scintigraphy and computed tomography. **J Clin Endocrinol Metab 2005**;90(6):3392-400.
49. Grossman AB, Kelly P, Rockall A, Bhattacharya S, McNicol A, Barwick T. Cushing's syndrome caused by an occult source: difficulties in diagnosis and management. **Nat Clin Pract Endocrinol Metab 2006**;2(11):642-7.

Address for correspondence:

John Newell-Price
Senior Lecturer and Consultant Endocrinologist
Academic Unit of Diabetes, Endocrinology & Metabolism
The University of Sheffield
Room OU142
O Floor
Royal Hallamshire Hospital
Glossop Road
Sheffield, S10 2JF, UK

A default mode of brain function

Marcus E. Raichle*†, Ann Mary MacLeod*, Abraham Z. Snyder*, William J. Powers‡, Debra A. Gusnard*§, and Gordon L. Shulman‡

*Mallinckrodt Institute of Radiology and Departments of ‡Neurology and §Psychiatry, Washington University School of Medicine, St. Louis, MO 63110

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected on April 30, 1996.

Contributed by Marcus E. Raichle, October 26, 2000

A baseline or control state is fundamental to the understanding of most complex systems. Defining a baseline state in the human brain, arguably our most complex system, poses a particular challenge. Many suspect that left unconstrained, its activity will vary unpredictably. Despite this prediction we identify a baseline state of the normal adult human brain in terms of the brain oxygen extraction fraction or OEF. The OEF is defined as the ratio of oxygen used by the brain to oxygen delivered by flowing blood and is remarkably uniform in the awake but resting state (e.g., lying quietly with eyes closed). Local deviations in the OEF represent the physiological basis of signals of changes in neuronal activity obtained with functional MRI during a wide variety of human behaviors. We used quantitative metabolic and circulatory measurements from positron-emission tomography to obtain the OEF regionally throughout the brain. Areas of activation were conspicuous by their absence. All significant deviations from the mean hemisphere OEF were increases, signifying deactivations, and resided almost exclusively in the visual system. Defining the baseline state of an area in this manner attaches meaning to a group of areas that consistently exhibit decreases from this baseline, during a wide variety of goal-directed behaviors monitored with positron-emission tomography and functional MRI. These decreases suggest the existence of an organized, baseline default mode of brain function that is suspended during specific goal-directed behaviors.

Functional brain imaging studies in normal human subjects with positron-emission tomography (PET) and functional MRI (fMRI) have consistently revealed expected task-induced increases in regional brain activity during goal-directed behaviors (for brief reviews see refs. 1 and 2). These changes are detected when comparisons are made between a task state, designed to place demands on the brain, and a control state, with a set of demands that are uniquely different from those of the task state.

Researchers have also frequently encountered task-induced decreases in regional brain activity even when the control state consists of lying quietly with eyes closed or passively viewing a stimulus. Whereas cortical increases in activity have been shown to be task specific and, therefore, vary in location depending on task demands, many decreases (Fig. 1) appear to be largely task independent, varying little in their location across a wide range of tasks (3). This consistency with which certain areas of the brain participate in these decreases made us wonder whether there might be an organized mode of brain function that is present as a baseline or default state and is suspended during specific goal-directed behaviors.

The primary issue this paper will address is whether these unexplained decreases merely arise from unrecognized increases (i.e., activation in the jargon of functional brain imaging) present only in the "control state." Thus, on this argument, any control state, no matter how carefully it is selected, is just another task state with its own unique areas of activation. Unfortunately, in most instances there is insufficient information about the control state to judge whether the observed decrease arose in this manner.

We believe conceptual progress has suffered because of our inability to exclude explanations of the above type for regional decreases in brain activity and, more generally, to understand whether a specific level of activity in a given area of the brain can be considered its baseline. At the heart of the problem is the lack of agreed-upon characteristics defining a baseline state. In response to this dilemma we began with a generally accepted, quantitative circulatory and metabolic definition of brain activation (see *Background*, below). From this definition we specified criteria for a baseline state (i.e., the absence of activation by this definition). In so doing, we were able to determine that areas consistently exhibiting decreases in activity during specific goal-directed behaviors (3) did so from this baseline state. We believe these findings are consistent with our idea of a baseline or default state of the brain, the functions of which are revealed by those areas whose activities are suspended during many transient, attention-demanding, goal-directed activities.

Background

Although the human brain accounts for only about 2% of the body weight, it consumes nearly 20% of the oxygen we extract from the air we breathe. This dependence of the brain on oxygen is highlighted by the fact that failure of oxygen delivery to the brain, usually the result of a stoppage of the heart, results in unconsciousness within seconds. An examination of the relationship between oxygen delivery to the brain and flowing blood regionally within the brain (Fig. 2) highlights the nature of this dependency.

The signal used by PET to map changes in neural activity in the human brain is based on local changes in blood flow (1). Increased neural activity in a local brain region increases blood flow in that region. Scientists have known of this robust relationship for well over 100 years through repeated demonstrations in laboratory animals and humans (1). It was thought to reflect the changing needs of the brain for oxygen during changing mental activity.

Surprisingly, more recently it has been appreciated that these changes in blood flow are accompanied by smaller changes in oxygen consumption (4). This leads to a decrease in the amount of oxygen extracted from blood when blood flow increases and an increase in the amount of oxygen extracted when blood flow decreases. Thus, changes in blood flow accompanying local changes in brain activity are associated with significantly smaller changes in the amount of oxygen used by the brain (1). As a result of these relationships the local blood oxygen content parallels the change in brain activity because the amount of oxygen supplied changes more than the demand (Fig. 3). At the present time we do not fully understand why the relationship between

Abbreviations: OEF, oxygen extraction fraction; fMRI, functional MRI; PET, positron-emission tomography; CBF, cerebral blood flow; $CMRO_2$, cerebral metabolic rate for oxygen; MPFC, medial prefrontal cortex.

†To whom reprint requests should be addressed at: Washington University School of Medicine, 4525 Scott Avenue, Room 2116, St. Louis, MO 63110. E-mail: marc@npg.wustl.edu.

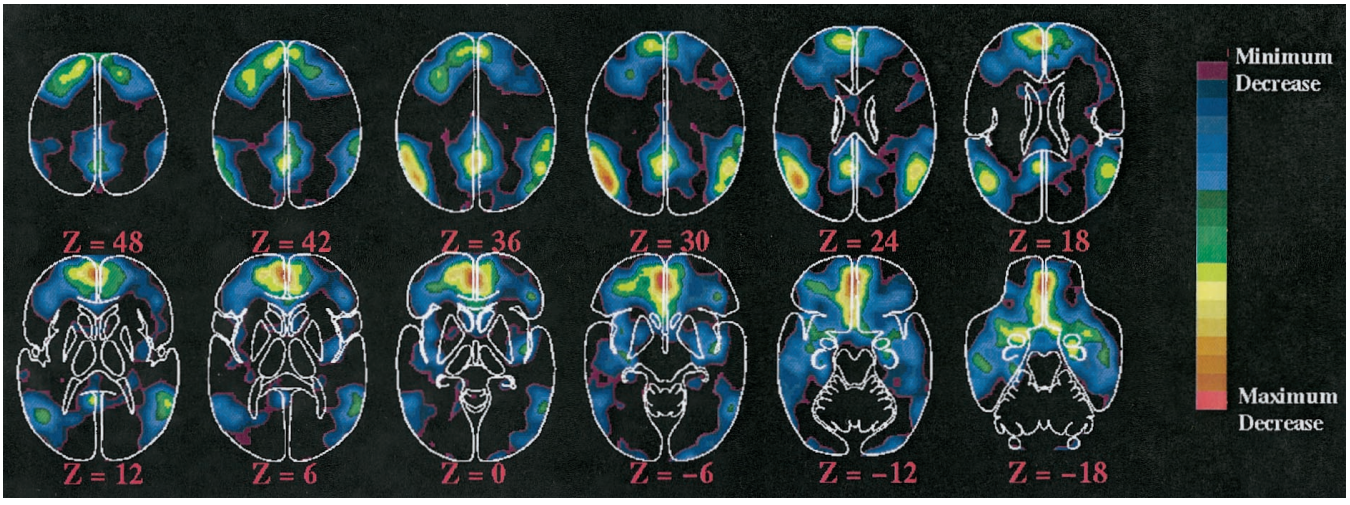


Fig. 1. Regions of the brain regularly observed to decrease their activity during attention demanding cognitive tasks. These data represent a metaanalysis of nine functional brain imaging studies performed with PET and analyzed by Shulman and colleagues (49). In each of the studies included, the subjects processed a particular visual image in the task state and viewed it passively in the control state. One hundred thirty-two individuals contributed to the data in these images. These decreases appear to be largely task independent. The images are oriented with the anterior at the top and the left side to the reader's left. The numbers beneath each image represent the millimeters above or below a transverse plane running through the anterior and posterior commissures (26).

oxygen delivery and oxygen consumption changes during changes in brain activity (see ref. 1 for a review), but this phenomenon has had great practical value for our ability to view changes in brain activity with fMRI.

Because fMRI signal intensity is sensitive to the amount of oxygen carried by hemoglobin (5–7), this change in blood oxygen content at the site of increased brain activity can be detected with fMRI. This phenomenon is the basis for fMRI (8, 9) and is usually referred to as the blood oxygen level-dependent (BOLD) signal, following Ogawa and colleagues (6).

The relationship of oxygen delivery to oxygen utilization can be measured quantitatively in the human brain with PET as the fraction of available oxygen (i.e., the arterial oxygen concentration) used by the brain. This measurement is usually referred to as the oxygen extraction fraction (OEF) (10–12). Researchers interested in blood flow and metabolic relationships in the brain have come to appreciate the spatial uniformity of the OEF measured in a resting state (e.g., lying quietly in a scanner with eyes closed but awake; see Fig. 4). This spatial uniformity exists despite considerable variation in resting oxygen consumption and blood flow within gray matter and an almost 4-fold difference between gray and white matter in both oxygen consumption and blood flow (Figs. 2 and 4). This relationship is altered in the normal brain only when areas briefly change their activity during specific behaviors (4, 13).

Heretofore the uniformity of the OEF at rest has not been considered in defining a baseline state of the human brain. Here we specifically propose to do so. The brain mean OEF was chosen as the baseline level of activity on the basis of its general uniformity in the eyes closed, resting state. This uniformity suggests that equilibrium has been reached between the local metabolic requirements necessary to sustain a long-term modal

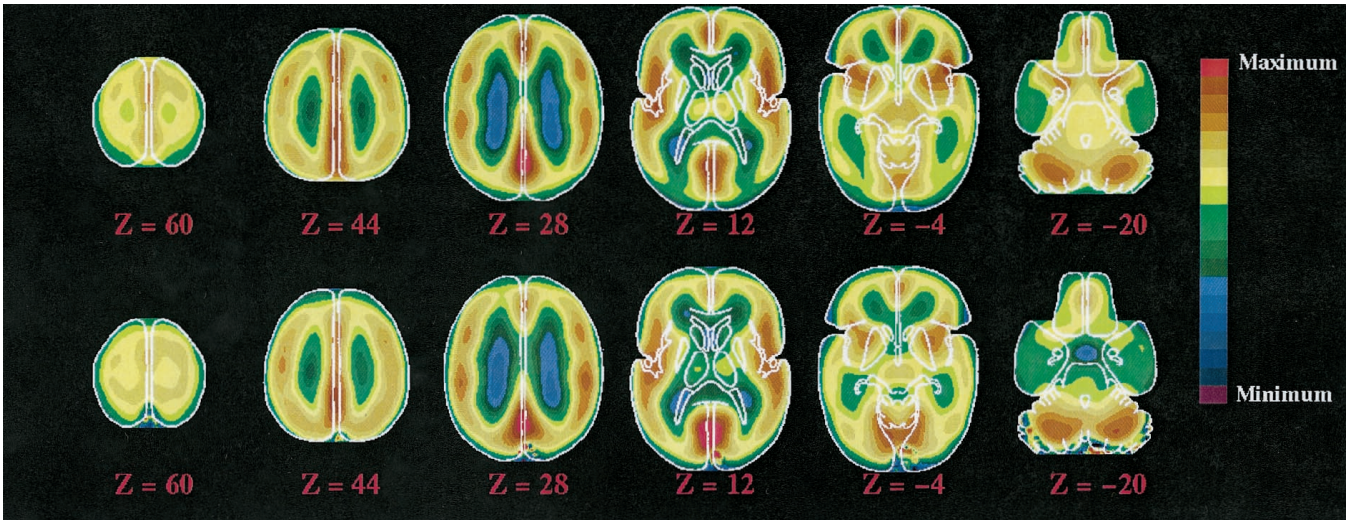


Fig. 2. Quantitative maps of blood flow (*Upper*) and oxygen consumption (*Lower*) in the subjects from group I while they rested quietly but awake with their eyes closed. The quantitative hemisphere mean values for these images are presented in Table 1. Note the large variation in blood flow and oxygen consumption across regions of the brain. These vary most widely between gray and white matter. Despite this variation, blood flow and oxygen consumption are closely matched, as also reflected in the image of the oxygen extraction fraction (i.e., the ratio of oxygen consumption to blood flow; see Fig. 4).

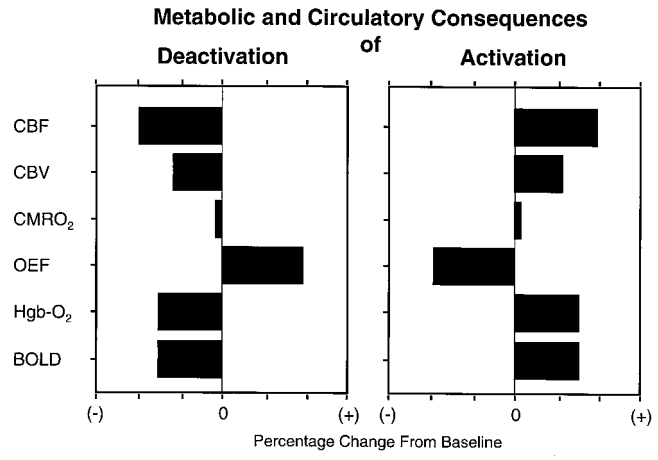


Fig. 3. A schematic representation of the metabolic and circulatory relationships occurring in areas of the brain with transient increases (Activation) or decreases (Deactivation) in the level of neural activity from a baseline or equilibrium state. Typically increases (*Right*) are characterized by increases in the cerebral blood flow (CBF) and the cerebral blood volume (CBV), with much smaller changes in the cerebral metabolic rate for oxygen ($CMRO_2$). As a result, there is a fall in the oxygen extraction fraction (OEF) and an increase in the amount of oxygen attached to hemoglobin exiting the brain (HbO_2). This latter change is responsible for the blood oxygen level–dependent (BOLD) signal used in functional magnetic resonance imaging (fMRI). Decreases from baseline (*Left*) are characterized as the opposite pattern of change.

level of neural activity and the level of blood flow in that region. We propose that this equilibrium state defines a baseline level of local neuronal activity. Consequently, those areas with a reduced OEF relative to the brain mean are defined as activated (i.e., neural activity is increased above the baseline level). Those areas not differing from the brain mean OEF are considered to be at baseline. In this scheme, increases in the OEF from the brain mean then define areas of deactivation (i.e., neural activity is decreased below the baseline level).

With these definitions in mind (Fig. 3) we used quantitative measurements of the OEF obtained with PET to examine those brain areas regularly observed to exhibit reductions in blood flow and blood oxygen level-dependent signal during goal-directed behaviors (3). Our reason for choosing these regions was to test the hypothesis that such decreases occur relative to a baseline state of brain activity (here defined as resting quietly but awake with eyes closed). In other words, decreases in brain activity do not have to be increases simply looked at from the opposite side of the equation. Rather, they are decreases from a true baseline or zero set point. If this hypothesis is correct, the chosen regions should exhibit an OEF similar to that of the rest of the brain in this baseline state. The OEF in these areas was measured in two independent groups of 19 normal adults resting quietly with eyes closed. These results were then extrapolated to a more complex control or baseline state (passive visual fixation) in a third group of 11 normal adults.

Methods

Subjects. Data from three subject groups were used for this analysis. The first two groups had served as control subjects in previously published studies from this laboratory (3, 14, 15). The third group was culled from previously unpublished data. All subjects were recruited from the Washington University community. None had any history of neurological or psychiatric illness. The Human Studies Committee and the Radioactive Drug Research Committee of Washington University approved all studies. Informed consent was obtained in accordance with their guidelines.

Group I consisted of 19 subjects (eight females) in whom quantitative CBF, CBV, OEF, and the $CMRO_2$ were measured while the subjects rested quietly but awake with eyes closed in a Siemens model 961 PET scanner (see below). Their ages ranged from 19 to 77 years, with a mean age of 43 years. Seventeen were right handed. They are described in greater detail elsewhere (15).

Group II consisted of 19 subjects (10 females) in whom quantitative CBF, CBV, OEF, and $CMRO_2$ were measured while the subjects rested quietly but awake with eyes closed in the PETT VI PET scanner (see below). Their ages ranged from 19 to 84 years, with a mean age of 42 years. Fourteen were right handed. They are described in greater detail elsewhere (14).

Group III (previously unpublished) consisted of 11 subjects

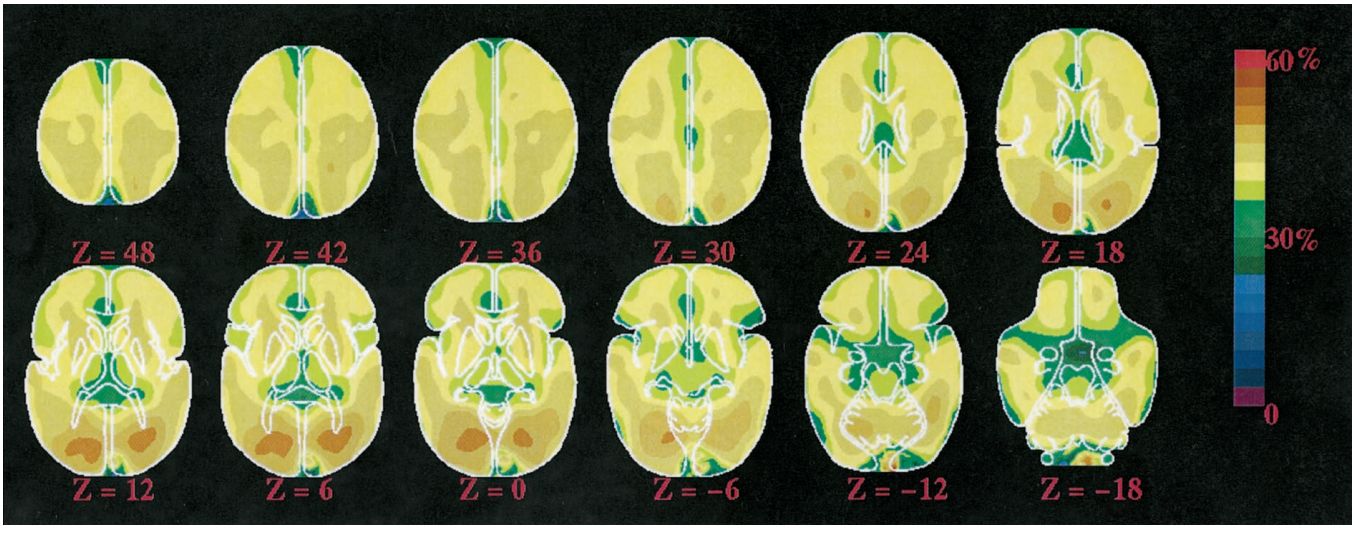


Fig. 4. Maps of the fraction of oxygen extracted by the brain from arterial blood (oxygen extraction fraction or OEF expressed as a percentage of the available oxygen delivered to the brain). The data come from 19 normal adults (group I, Table 1) resting quietly but awake with their eyes closed. The data were obtained with PET. Despite an almost 4-fold difference in blood flow and oxygen consumption between gray and white matter, the OEF is relatively uniform, emphasizing the close matching of blood flow and oxygen consumption in the resting, awake brain. Areas of increased OEF can be seen in the occipital regions bilaterally (see text for discussion).

Table 1. Data obtained from subjects in group I while they rested quietly but awake with their eyes closed

	Area	*x*	*y*	*z*	OEF	*P*	CBF	*P*	$CMRO_2$	*P*
A	M 31/7	−5	−49	40	1.010	0.62	1.374*	<0.0001	1.397*	<0.0001
B	L 40	−53	−39	42	0.901	0.07	0.735*	<0.0001	0.695*	<0.0001
C	L 39/19	−45	−67	36	0.987	0.66	0.813*	0.0008	0.805*	0.0003
D	R 40	45	−57	34	0.978	0.26	1.002	0.96	0.998	0.97
E	L lateral 8	−27	27	40	0.981	0.52	0.991	0.74	0.986	0.74
F	L 8/9	−11	41	42	0.902	0.04	0.874	0.006	0.803*	0.001
G	R 8/9	5	49	36	0.887	0.004	0.893	0.01	0.813*	0.001
H	L 9	−15	55	26	0.943	0.35	0.813*	0.0001	0.785*	0.002
I	L 10	−19	57	8	0.971	0.06	0.940	0.01	0.927	0.01
J	M 10	−1	47	−4	0.933	0.01	1.284*	<0.0001	1.203*	<0.0001
K	L 10/47	−33	45	−6	0.930	0.25	0.920	0.03	0.879	0.04
L	M 32	3	31	−10	0.946	0.20	1.111	0.005	1.058	0.26
M	L 20	−49	−19	−18	0.972	0.46	0.821*	<0.0001	0.814*	0.0002

The values for OEF, CBF, and $CMRO_2$ are expressed as local-to-global ratios (see *Methods*). For the 19 subjects in group I the global values for OEF and CBF (± SD) were 0.40 ± 0.09 (dimensionless) and 46 ± 8 ml/(min × 100 *g*), respectively. The mean arterial oxygen content for this group was 16.6 ± 0.16 ml/ml. From these data quantitative images of the $CMRO_2$ were created that yielded a mean cerebral hemisphere value of 2.94 ± 0.41 ml/(min × 100 *g*) or 1.31 ± 0.18 μmol/(min × *g*). The asterisks denote values that differ significantly from the global mean after correction for multiple comparisons.

(five females) in whom qualitative CBF was measured while the subjects rested quietly but awake with eyes closed and again while they passively viewed a visual fixation cross hair in the middle of a television monitor. Five such paired measurements were obtained. Their ages ranged from 19 to 40 years, with a mean age of 27 years. Eight subjects were right handed. There is no record of the handedness of the other three. The data for group III were also acquired with the PETT VI PET scanner (see below).

Imaging Methods. The subjects in group I were scanned with a Siemens model 961 scanner (Siemens Medical Systems, Hoffman Estates, IL) (16, 17). Images were reconstructed by using filtered back projection and scatter correction with a ramp filter at the Nyquist frequency. All images were then filtered with a three-dimensional Gaussian filter to a uniform resolution of 16-mm full width at half-maximum.

In the subjects in groups II and III, studies were performed on the PETT VI scanner (18, 19). The PETT VI system was used in the low-resolution mode. Images were then filtered with a three-dimensional Gaussian filter to a uniform resolution of 17-mm full width at half-maximum.

Quantitative regional OEF, CBF, and $CMRO_2$ were measured by using a combination of O^{15}-labeled radiopharmaceuticals as described (11, 12, 20–22). Qualitative CBF was measured in the subjects in group III. In this instance $H_2{}^{15}O$ images consisting of normalized PET counts were used to create maps of the distribution of blood flow in the brain (21, 23, 24).

Regions of Interest. The location of each region of interest analyzed in this study was obtained directly from the right hand side of table 1 of the study by Shulman and colleagues (3). These regions were chosen because they reliably predict the location of areas of the human brain exhibiting reductions in activity, as measured with either PET or fMRI, during the performance of a variety of cognitive tasks (3). Their coordinates are listed in Tables 1 and 2 of the present study. The regions are referred to by the Brodmann area originally assigned to them by Shulman *et al.* (3).

Furthermore, a center-of-mass algorithm (25) was used to search the OEF data sets for any significant deviations from the hemisphere mean outside of the specific areas chosen for analysis (preceding paragraph).

Data Analysis. A 12-mm sphere was centered on the stereotaxic coordinates of each region of interest (see Tables 1 and 2) for each subject's individual images of CBF, OEF, and $CMRO_2$. Each regional value was divided by the subject's mean global

Table 2. Data obtained from subjects in group II while they rested quietly but awake with their eyes closed

	Area	*x*	*y*	*z*	OEF	*P*	CBF	*P*	$CMRO_2$	*P*
A	M 31/7	−5	−49	40	1.035	0.62	1.278*	<0.0001	1.345*	<0.0001
B	L 40	−53	−39	42	0.903	0.04	0.804	0.0042	0.756*	0.003
C	L 39/19	−45	−67	36	0.997	0.94	0.805*	<0.0001	0.820*	0.0004
D	R 40	45	−57	34	0.994	0.77	0.941	0.04	0.957	0.11
E	L lateral 8	−27	27	40	1.001	0.93	0.964	0.05	0.996	0.93
F	L 8/9	−11	41	42	0.707	0.04	0.893	0.03	0.756	0.02
G	R 8/9	5	49	36	0.871	0.02	0.926	0.15	0.858	0.02
H	L 9	−15	55	26	0.983	0.54	0.876*	0.0002	0.878*	0.0009
I	L 10	−19	57	8	1.042	0.06	0.967	0.02	1.012	0.61
J	M 10	−1	47	−4	0.984	0.36	1.187*	<0.0001	1.166*	<0.0001
K	L 10/47	−33	45	−6	1.037	0.12	0.913*	<0.0001	0.964	0.27
L	M 32	3	31	−10	0.984	0.54	1.080*	<0.0001	1.061	0.03
M	L 20	−49	−19	−18	0.962	0.35	0.865*	<0.0001	0.877	0.005

The values for OEF, CBF, and $CMRO_2$ are expressed as local-to-global ratios (see *Methods*). For the 19 subjects in group II the global values for OEF and CBF (± SD) were 0.30 ± 0.09 (dimensionless) and 48 ± 10 ml/(min × 100 *g*), respectively. The mean arterial oxygen content for this group was 17.0 ± 0.02 ml/ml. From these data quantitative images of the $CMRO_2$ were created that yielded a mean cerebral hemisphere value of 2.17 ± 0.41 ml/(min × 100 *g*) or 0.97 ± 0.17 μmol/(min × *g*). The asterisks denote values that differ significantly from the global mean after correction for multiple comparisons.

Table 3. Areas with maximum absolute deviation of the OEF from the hemisphere mean in subjects from group I

Brodmann area	*x*	*y*	*z*	OEF	*P*
11	8	33	−22	1.491	0.0007
19	−23	−79	6	1.217	<0.0001
19	29	−67	4	1.201	<0.0001
18	−29	−93	12	1.194	0.001
18	23	−79	18	1.182	0.0002
18	23	−81	10	1.181	<0.0001
37	33	−51	−8	1.138	0.0005
31	23	−47	40	1.121	<0.0001
31	17	−33	50	1.103	0.0003
19	21	−83	30	1.098	0.08

Note that these are all increases in the OEF signifying areas of deactivation in the baseline state of resting but awake with eyes closed. With the exception of Brodmann area 11 located in the right gyrus rectus, these areas cluster in visual areas of the occipital and parietal cortices. The conventions used are similar to those of Tables 1 and 2.

value for that image, creating a local to global ratio for that region. A two-tailed, one-sample Student's *t* test was performed to determine whether this ratio differed from a predicted value of 1. The level of significance was adjusted for multiple comparisons.

Results

Regional as well as whole-brain measurement values of CBF, OEF, and $CMRO_2$ for groups I and II are presented in Tables 1 and 2. None of the areas selected for study (3) exhibited an OEF significantly different from the hemispheric mean (compare Figs. 1 and 4). The correlation between the regional values of OEF obtained independently on the two groups was excellent ($r = 0.89$). By our definition (see *Background*), therefore, none of these areas was activated in subjects who were awake but resting quietly with their eyes closed (two independent groups of 19 subjects each).

To complete our analysis, the data sets from groups I and II were automatically searched (25) for any deviations in the OEF above or below the hemisphere mean. No significant decreases in OEF were found signifying areas of activation. In arriving at this determination we had to exclude several "areas" falling in high-noise areas at the edge of the brain or, in some instances, clearly outside of the brain. However, we did find, bilaterally, areas within extrastriate visual cortices that exhibited a significantly increased OEF from the hemisphere mean. These changes are readily apparent in Fig. 4, and their locations are given in Tables 3 and 4. By our definition these apparent visual areas are deactivated when subjects are awake but resting quietly with their eyes closed.

Table 4. Areas with maximum absolute deviation of the OEF from the hemisphere mean in subjects from group II

Brodmann area	*x*	*y*	*z*	OEF	*P*
18	19	−79	10	1.189	<0.0001
18	−19	−93	14	1.169	<0.0001
19	35	−75	10	1.163	<0.0001
19	−27	−79	8	1.145	0.001
17	−9	−75	6	1.123	0.003
19	21	−81	28	1.112	0.03
18	−13	−69	30	1.109	0.002
7	21	−59	32	1.107	<0.0001
19	−23	−87	28	1.104	.05
7	17	−63	46	1.073	0.02

Note that these are all increases in the OEF signifying areas of deactivation in the baseline state of resting but awake with eyes closed. These areas cluster in visual areas of occipital and parietal cortices. The conventions used are similar to those of Tables 1 and 2.

Although none of the areas selected for study appear to be activated by our definition, two areas in both groups (i.e., M31/7 and M10) have both resting CBF and $CMRO_2$ significantly above the global mean (criteria: $P < 0.0038$ for both measurements in both groups). Likewise, two additional areas in both groups (i.e., L39/19 and L9) have both resting CBF and $CMRO_2$ significantly below the global mean.

Resting quietly awake with the eyes closed is much less frequently used as a control state in functional imaging studies than visual fixation of a cross hair on a television monitor or passive viewing of a stimulus. The data presented in Tables 1 and 2 as well as Figs. 1 and 4 do not tell us whether areas exhibiting significant reductions in activity during active task performance (3) become more active in the control states of visual fixation than when the eyes are closed. If this were to occur, reduced activity seen in the task state could simply reflect the absence of this increase in activity during fixation and passive viewing.

The data from group III resolved this issue. No significant change in blood flow was found for any of the regions listed in Tables 1 and 2 when subjects went from the eyes closed and awake but resting state to passively viewing a fixation cross hair in the middle of a television monitor. Furthermore, it should be noted that those areas of extrastriate visual cortex exhibiting deactivation in the eyes closed state (Fig. 4) increase their BF when the eyes are opened (details to be published separately). This observation is consistent with the hypothesis that the baseline state of these areas is more nearly approximated when subjects rest quietly with their eyes open.

Discussion

Our study represents a comprehensive analysis of the uniformity of the OEF in the normal human brain while adult subjects are awake and resting quietly with their eyes closed. These data affirm the long-held impression of a relatively consistent relationship between oxygen delivery and oxygen consumption in the human brain (Fig. 2). Obvious decreases in the OEF from the brain mean, reflecting areas of activation (1), are not apparent in our data when subjects rest quietly with their eyes closed or open.

Areas of deactivation (i.e., increased OEF), primarily in extrastriate visual areas, were clearly apparent from our data (Fig. 4 and Tables 3 and 4). It is of interest to note that these same increases in OEF were also noted in some of the earliest PET work on normal humans (27), although their possible significance was not appreciated. Their presence suggests that the baseline state for these areas may well be associated with the eyes being open. This hypothesis receives support from our comparison of eyes closed versus passive visual fixation, in which we noted an increase in the blood flow in these areas as subjects opened their eyes. Further work remains, including actual measurements of OEF in these areas in the eyes open and closed states. Nevertheless, these data are consistent with our hypothesis that the baseline state of these areas is more nearly approximated when the eyes are open.

More generally our data are consistent with the hypothesis that the brain mean OEF defines a baseline level of neuronal activity. The uniformity of the OEF in the absence of specific goal-directed activities supports our belief that an established equilibrium exists between the local metabolic requirements necessary to sustain a long-term modal level of neural activity and the level of blood flow in a particular region. Deviations from this equilibrium produced by transient changes in this modal level of neural activity manifest themselves as changes in the OEF and provide us with the signals underlying modern functional brain imaging with fMRI.

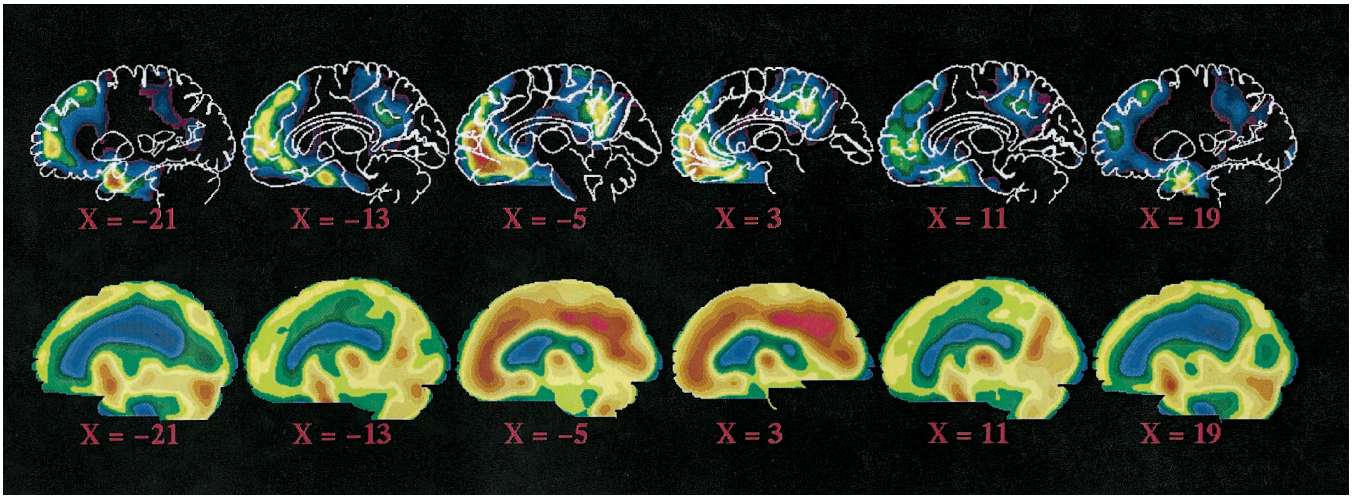


Fig. 5. Regions of the brain regularly observed to decrease their activity during attention-demanding cognitive tasks shown in sagittal projection (*Upper*) as compared with the blood flow of the brain while the subject rests quietly but is awake with eyes closed (*Lower*). The data in the top row are the same as those shown in Fig. 1, except in the sagittal projection, to emphasize the changes along the midline of the hemispheres. The data in the bottom row represent the blood flow of the brain and are the same data shown in horizontal projection in the top row of Fig. 2. The numbers below the images refer to the millimeters to the right (positive) or left (negative) of the midline.

Areas consistently observed to decrease their activity in a variety of goal-directed, cognitive activation paradigms (Fig. 1) do not exhibit a reduced OEF (i.e., evidence of activation) in this typical resting state (Tables 1 and 2). While a null result such as this (i.e., the regional OEF does not differ from the hemisphere mean) might be received with some caution, the relatively tight 95% confidence limits for the measurement of OEF (±3%, based on an analysis of the group I data) make it very unlikely that significant reductions in OEF, of even lesser magnitude than the increases seen in the visual system (Fig. 4 and Tables 3 and 4), would have been missed. Thus, we believe our findings indicate that these localized decreases in activity (Figs. 1 and 5) (3) occur as decreases from a baseline level of activity rather than the return to baseline of an area of unsuspected activation.

The presence of a consistent set of decreases in activity within a select set of brain areas strikingly independent of the goal-directed behaviors with which they are associated suggests to us that specific brain functions unique to the baseline state itself are being temporarily suspended. We posit that areas decreasing their activity in this manner may be tonically active in the baseline state, as distinguished from areas that are transiently activated in support of varying goal-directed activities. Understanding the exact functions served by such tonically active areas would require much additional work, yet some indication of the directions this research might profitably take is revealed by our current knowledge about two of them. These are midline areas within the posterior cingulate and precuneus and within the medial prefrontal cortex (MPFC) (see Fig. 5).

The response of posterior cingulate and precuneus neurons to visual stimuli, for example, crucially depends on the physical characteristics of the stimulus (for a recent review see ref. 28). Small spots of light to which a monkey may be attending and responding do not elicit neuronal responses in this area. In contrast, large, brightly textured stimuli elicit responses, even if they are totally irrelevant to tasks the animal is performing. These observations are consistent with the fact that elements of the dorsal stream of extrastriate visual cortex [area M in the owl monkey (29) and area PO in the macaque (30)] are part of a network of areas concerned with the representation of the visual periphery.

Severe damage to the parietal cortex, when it extends medially to include the precuneus and the posterior cingulate, produces a condition known as Balint's syndrome (31), the cardinal feature of which is the inability to perceive the visual field as a whole (i.e., a fixed form of tunnel vision usually referred to as simultanagnosia), despite intact visual fields, during simple confrontation with single small stimuli. Furthermore, patients with Alzheimer's disease show early reductions of metabolic activity in this area (32) and have been reported to show abnormalities in the processing of extrafoveal information (33).

Thus, the posterior cingulate cortex and adjacent precuneus can be posited as a tonically active region of the brain that may continuously gather information about the world around, and possibly within, us. It would appear to be a default activity of the brain with rather obvious evolutionary significance. Detection of predators, for example, should not, in the first instance, require the intentional allocation of attentional resources. These resources should be allocated automatically and be continuously available. Only when successful task performance demands focused attention should such a broad information gathering activity be curtailed. The central importance of such a function is underscored by the observation that restoration of consciousness from a vegetative state (or at least external awareness, as it can be assessed at the patient's bedside) is primarily heralded by a restoration of metabolism in parietal areas, including the precuneus (34).

Finally, we note the selective vulnerability of the posterior cingulate and precuneus in such conditions as carbon monoxide poisoning [i.e., acute hypoxia (34)], diffuse brain ischemia (35), and Alzheimer's disease (32). This vulnerability, in the case of hypoxia and ischemia, has been ascribed to the position of the posterior cingulate and precuneus in the border zone between two of the main arteries supplying blood to the brain. We wonder whether the exceptionally high metabolic rate exhibited by the posterior cingulate and precuneus adds to their vulnerability (Tables 1 and 2 and Fig. 5). This hypothesis receives some support from animal models of schizophrenia (36), based on pharmacologically induced, excitatory amino acid toxicity that preferentially (but without explanation to date) targets this area.

Another midline area of the cortex exhibiting prominent decreases in activity during focused attention is MPFC (Figs. 1 and 5). Because of the large body of data implicating, in particular, the ventral aspects of this area of the brain and its connections in emotional processing within the brain, it has been

posited that decreases during focused attention reflect a dynamic interplay between ongoing cognitive processes and the emotional state of the subject (37–40), a hypothesis we explore in more detail in our accompanying papers (41, 42). More specifically, it has been proposed that in humans the MPFC may play a role in the integration of emotional and cognitive processes by incorporating emotional biasing signals or markers into decision-making processes (43).

Anatomically, the ventral MPFC is composed of a number of cytoarchitectonically discrete areas that receive a wide range of sensory information from the body and the external environment (44–46) and are heavily interconnected with limbic structures such as the amygdala, ventral striatum, hypothalamus, midbrain periaquaductal gray region, and brainstem autonomic nuclei (44, 45). Electrophysiological studies in monkeys indicate that orbital and medial prefrontal neurons respond to stimuli based on their reward and nonreward associations (for a review see ref. 47).

Thus, as we come to associate general monitoring of sensory information with the posterior cingulate and adjacent precuneus, we may also come to associate an evaluation of the salience of this information with medial and orbital frontal cortices. We posit that when an individual is awake and alert and yet not actively engaged in an attention-demanding task, a default state of brain activity exists that involves, among other areas, the MPFC and the posterior cingulate and precuneus. Information broadly arising in the external and internal milieu is gathered and evaluated. When focused attention is required, particularly if this activity is novel, activity within these areas may be attenuated. This attenuation in activity reflects a necessary reduction in resources devoted to general information gathering and evaluation. The MPFC with the posterior cingulate and medial parietal cortices may well be the "sentinels" to which William James referred (see ref. 48, p. 73), which, "when beams of light move over them, cry 'who goes there' and call the fovea to the spot. Most parts of the skin do but perform the same office for the fingertips."

More generally, an appreciation of the important activities that may underlie the baseline state of the human brain will certainly enrich our understanding of its function. Strong hints about the direction research in this area might take come, we believe, from the heretofore mysterious activity reductions routinely seen by modern functional brain imaging techniques during the performance of goal-directed behaviors. In addition to the midline areas we have discussed, it is necessary to consider as well the role(s) of more lateral cortical areas (especially prominent in parietal cortex), which also appear as elements of this default system.

This work was supported by National Institutes of Health Grants NS06833, DA07261, and NS10196 and the Charles A. Dana Foundation.

1. Raichle, M. E. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 765–772.
2. Raichle, M. E. (1998) *Philos. Trans. R. Soc. London B* **353,** 1–14.
3. Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E. & Petersen, S. E. (1997) *J. Cognit. Neurosci.* **9,** 648–663.
4. Fox, P. T. & Raichle, M. E. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1140–1144.
5. Thulborn, K. R., Waterton, J. C., Matthews, P. M. & Radda, G. K. (1982) *Biochim. Biophys. Acta* **714,** 265–270.
6. Ogawa, S., Lee, T. M., Kay, A. R. & Tank, D. W. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 9868–9872.
7. Ogawa, S., Lee, T. M., Nayak, A. S. & Glynn, P. (1990) *Magn. Reson. Med.* **14,** 68–78.
8. Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S.-G., Merkle, H. & Ugurbil, K. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5951–5955.
9. Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weiskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., *et al.* (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5675–5679.
10. Frackowiak, R. S. J., G. L., L., Jones, T. & Heather, J. D. (1980) *J. Comput. Assist. Tomogr.* **4,** 727–736.
11. Mintun, M. A., Raichle, M. E., Martin, W. R. & Herscovitch, P. (1984) *J. Nucl. Med.* **25,** 177–187.
12. Altman, D. I., Lich, L. L. & Powers, W. J. (1991) *J. Nucl. Med.* **32,** 1738–1741.
13. Fox, P. T., Raichle, M. E., Mintun, M. A. & Dence, C. (1988) *Science* **241,** 462–464.
14. Perlmutter, J. S., Herscovitch, P., Powers, W. J., Fox, P. T. & Raichle, M. E. (1985) *J. Cereb. Blood Flow Metab.* **5,** 476–480.
15. Grubb, R. L., Derdeyn, C. P., Fritsch, S. M., Carpenter, D. A., Yundt, K. D., Videen, T. O., Spitznagel, E. L. & Powers, W. J. (1998) *J. Am. Med. Assoc.* **280,** 1055–1060.
16. Wienhard, K., M., D., L., E., Michel, C., Bruckbauer, T., Pietrzyk, U. & Heiss, W. D. (1994) *J. Comput. Assist. Tomogr.* **18,** 110–118.
17. Spinks, T. J., Jones, T., Bailey, D. L., Townsend, D. W., Grootoonk, S., Bloomfield, P. M., Gilardi, M. C., Casey, M. E., Sipe, B. & Reed, J. (1992) *Phys. Med. Biol.* **37,** 1637–1655.
18. Ter-Pogossian, M. M., Ficke, D. C., Hood, J. T., Yamamoto, M. & Mullani, N. A. (1982) *J. Comput. Assist. Tomogr.* **6,** 125–133.
19. Yamamoto, M., Ficke, D. C. & Ter-Pogossian, M. M. (1982) *IEEE Trans. Nucl. Sci.* **NS-29,** 529–533.
20. Videen, T. O., Perlmutter, J. S., Herscovitch, P. & Raichle, M. E. (1987) *J. Cereb. Blood Flow Metab.* **7,** 513–516.
21. Herscovitch, P., Markham, J. & Raichle, M. E. (1983) *J. Nucl. Med.* **24,** 782–789.
22. Raichle, M. E., Martin, W. R. W., Herscovitch, P., Mintun, M. A. & Markham, J. (1983) *J. Nucl. Med.* **24,** 790–798.
23. Fox, P. T. & Mintun, M. A. (1989) *J. Nucl. Med.* **30,** 141–149.
24. Fox, P. T., Mintun, M. A., Raichle, M. E. & Herscovitch, P. (1984) *J. Cereb. Blood Flow Metab.* **4,** 329–333.
25. Mintun, M. A., Fox, P. T. & Raichle, M. E. (1989) *J. Cereb. Blood Flow Metab.* **9,** 96–103.
26. Talairach, J. & Tournoux, P. (1988) *Co-Planar Stereotaxic Atlas of the Human Brain* (Thieme Medical Publishers, New York).
27. Lebrun-Grandie, P., Baron, J.-C., Soussaline, F., Loch'h, C., Sastre, J. & Bousser, M.-G. (1983) *Arch. Neurol.* **40,** 230–236.
28. Vogt, B. A., Finch, D. M. & olson, C. R. (1992) *Cereb. Cortex* **2,** 435–443.
29. Baker, J. F., Petersen, S. E., Newsome, W. T. & Allman, J. M. (1981) *J. Neurophysiol.* **45,** 397–416.
30. Colby, C. L., Gattass, R., Olson, C. R. & Gross, C. G. (1988) *J. Comp. Neurol.* **238,** 1257–1299.
31. Hecaen, H. & Ajuriaguerra, J. (1954) *Brain* **77,** 373–400.
32. Minoshima, S., Giordani, B., Berent, S., Frey, K. A., Foster, N. L. & Kuhl, D. E. (1997) *Ann. Neurol.* **42,** 85–94.
33. Benson, D. F., Davis, J. & Snyder, B. D. (1988) *Arch. Neurol.* **45,** 789–793.
34. Laureys, S., Lemaire, C., Maquet, P., Phillips, C. & Franck, G. (1999) *Lancet* **67,** 121–133.
35. DeVolder, A. G., Goffinet, A. M., Bol, A., Michel, C., de Barsy, T. & Laterre, C. (1990) *Arch. Neurol.* **47,** 197–204.
36. Olney, J. W., Newcomer, J. W. & Farber, N. B. (1999) *J. Psychiatric Res.* **33,** 523–533.
37. Drevets, W. C. & Raichle, M. E. (1998) *Cognit. Emotion* **12,** 353–385.
38. Mayberg, H. S. (1997) *J. Neuropsychiatry* **9,** 471–481.
39. Bush, G., Luu, P. & Posner, M. I. (2000) *Trends Cognit. Sci.* **4,** 215–222.
40. Simpson, J. R., Jr., Ongur, D., Akbudak, E., Conturo, T. E., Ollinger, J. M., Snyder, A. Z., Gusnard, D. A. & Raichle, M. E. (2000) *J. Cognit. Neurosci.*, in press.
41. Simpson, J. R., Jr., Snyder, A. Z., Gusnard, D. A. & Raichle, M. E. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 683–687.
42. Simpson, J. R., Jr., Drevets, W. C., Snyder, A. Z., Gusnard, D. A. & Raichle, M. E. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 688–693.
43. Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. (1997) *Science* **275,** 1293–1295.
44. Barbas, H. (1995) *Neurosci. Biobehav. Rev.* **19,** 499–510.
45. Carmichael, S. T. & Price, J. L. (1995) *J. Comp. Neurol.* **363,** 615–641.
46. Rolls, E. T. & Baylis, L. L. (1994) *J. Neurosci.* **14,** 5437–5452.
47. Rolls, E. T. (1999) *The Brain and Emotion* (Oxford Univ. Press, Oxford, U.K.).
48. James, W. (1893) *Psychology* (Henry Holt, New York).
49. Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E. & Petersen, S. E. (1997) *J. Cognit. Neurosci.* **9,** 648–663.

OPEN ACCESS Freely available online

PLOS ONE

Disgust and the Politics of Sex: Exposure to a Disgusting Odorant Increases Politically Conservative Views on Sex and Decreases Support for Gay Marriage

Thomas G. Adams[1], Patrick A. Stewart[2]*, John C. Blanchar[1]

1 Department of Psychological Science, University of Arkansas, Fayetteville, Arkansas, United States of America, **2** Department of Political Science, University of Arkansas, Fayetteville, Arkansas, United States of America

Abstract

Disgust has been implicated as a potential causal agent underlying socio-political attitudes and behaviors. Several recent studies have suggested that pathogen disgust may be a causal mechanism underlying social conservatism. However, the specificity of this effect is still in question. The present study tested the effects of disgust on a range of policy preferences to clarify whether disgust is generally implicated in political conservatism across public policy attitudes or is uniquely related to specific content domains. Self-reported socio-political attitudes were compared between participants in two experimental conditions: 1) an odorless control condition, and 2) a disgusting odor condition. In keeping with previous research, the present study showed that exposure to a disgusting odor increased endorsement of socially conservative attitudes related to sexuality. In particular, there was a strong and consistent link between induced disgust and less support for gay marriage.

Citation: Adams TG, Stewart PA, Blanchar JC (2014) Disgust and the Politics of Sex: Exposure to a Disgusting Odorant Increases Politically Conservative Views on Sex and Decreases Support for Gay Marriage. PLoS ONE 9(5): e95572. doi:10.1371/journal.pone.0095572

Editor: Brock Bastian, University of Queensland, Australia

Received July 10, 2013; **Accepted** March 27, 2014; **Published** May 5, 2014

Copyright: © 2014 Adams et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The funds used to pay research participants were provided by a small intramural award – the Connor Faculty Fellowship – from the Department of Political Science at the University of Arkansas, which was awarded to Dr. Patrick Stewart. The funders of this award (Sandra and Robert Connor) had no direct involvement in any aspect of the current study.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: pastewar@uark.edu

Introduction

Disgust is increasingly seen as playing an important role in the formation and maintenance of political and social attitudes [1]. Disgust sensitivity and reactivity have been shown to relate to political conservatism generally [2,3,3–7] and have been implicated in negative and hostile attitudes towards a range of social out-groups, particularly homosexuals [2,6–10]. In keeping with the aforementioned research, some have suggested that disgust is specifically related to conservative attitudes that pertain to social or moral purity [11], and, in particular, sexual purity [4,12,13]. One potential reason for these relations is that disgust plays an adaptive role in preserving individuals from the threat of infectious disease [14,15].

It has been posited that the Behavioral Immune System, which uses the physiological response of disgust as its first line of behavioral defense, shapes political attitudes and resultant policy preferences [5,16]. As stated by Terrizzi, Shook and McDaniel in their meta-analysis "(A)cross 24 studies, the findings suggest that there is a moderate, positive relationship between behavioral immune system strength and social conservatism" [7]. Despite connections being made between disgust and political ideology, additional work is needed to clarify if disgust sensitivity and/or disgust reactivity are associated with political ideology and what domains of political ideology are related to disgust. Smith and colleagues [2] recently addressed these questions by investigating the relations between disgust sensitivity, disgust reactivity – meaning *in vivo* physiological reaction to disgusting stimuli – and responses to a range of specific political issue items using the well-known Wilson-Patterson scale [17]. Heightened disgust sensitivity and reactivity were both negatively related to attitudes towards gay marriage and premarital sex but not significantly related to a range of other political issues (e.g., gun rights, school prayer, immigration). They also showed that, after controlling for demographic and experimental factors, greater disgust sensitivity was associated with less support for abortion. Said otherwise, this study showed that both trait and state disgust predicted less support for gay marriage and premarital sex and trait disgust predicted less support for abortion. It is noteworthy that, contrary to their predictions, Smith and colleagues failed to show any significant relation between state or trait disgust and attitudes about pornography. This is, however, not too surprising given the limited body of research showing a link between disgust and attitudes about pornography [18].

Evidence from Smith et al. [2] and similar research [9] is compelling but causal interpretations are not possible given their utilization of correlational designs. This limitation was rectified by Inbar and colleagues [3] who experimentally tested the effects of odor induced disgust on self-reported warmth toward homosexuals. They found that, relative to participants in a no odor control condition, participants in the disgust odorant condition reported less warmth toward homosexuals. These findings are similar to those reported by Terrizzi and colleagues [6], who showed that the induction of disgust caused conservative participants to report a decreased willingness to come into contact with homosexuals. What remains to be seen is if the experimental induction of disgust causes attitudinal shifts for other issues related to sexual purity

(e.g., premarital sex, abortion, etc.) and if induced disgust causes changes in attitudes towards public policies related to gay marriage.

In the present study we extend previous research by carrying out a true experiment to test the role of induced disgust on the same socio-political attitudes measured by Smith and colleagues. This allows us to test the causal role of disgust on a variety of public policy stances while still focusing on issues of sexual purity, particularly gay marriage. Core/contamination disgust was induced via noxious odorant, which is a primal approach closely reflecting the evolutionary adaptation of this emotion [19]. Moreover, manipulating physical disgust via smell allowed for a continuous induction of the emotion, rather than a brief priming of the emotion prior to measurement. In keeping with the findings of Smith and colleagues [2], we expected the foul disgust odorant, in this case butyric acid, which smells like vomit, to robustly engage the emotion of disgust and decrease approval of gay marriage. Smith and colleagues [2] reported weak effects between state disgust and attitudes about premarital sex and also found non significant relations between state disgust and attitudes about abortion or pornography. Similarly, both Smith and colleagues [2] and Inbar and colleagues [9] reported weak, albeit significant, relations between disgust sensitivity and attitudes about abortion. We therefore expected weak, or perhaps non-significant, effects of our experimental manipulation on attitudes about premarital sex, pornography, and abortion.

Methods

Procedures

All procedures were approved by the University of Arkansas Institutional Review Board. All participants provided written informed consent prior to participating in any experimental procedures. Data for the present study can be found at http://dx.doi.org/10.6084/m9.figshare.930730. Community members and students were recruited via an online advertisement and compensated $10 for their time. All procedures were completed in a small group format ($n = 4$–8). Following informed consent, participants were then escorted into a small room (10′x10′x17.5′) with individual desks. An experimenter operated a timed slideshow that displayed experimental instructions while monitoring participants at the back of the room, out of their sight. Participants were assigned to either a disgust odor condition or a control (no odor) condition. Ten minutes prior to participant arrival, an experimenter placed four drops of 99% butyric acid [20,21] on two small cotton pads and hid these pads at the front and back of the room. Pads were replaced and the room aerated between experimental groups. Immediately upon entering the experimental room, all participants completed disgust ratings and socio-political opinion questions. The Three Domains of Disgust Scale (TDDS), Conservative Liberal Scale, and demographics were completed in a separate, odor-neutral corridor (measures are listed in order of administration below). The disgust induction was not, therefore, expected to have any effect on the TDDS or Conservative Liberal Scale scores.

Participants

Smith and colleagues reported medium ($r = .37$) to large ($r = .45$) relations between disgust reactivity (state disgust) and attitudes about gay marriage [2]. Power analyses were conducted using G*Power [22]. Results suggest that, given alpha of .05 and power of .80, a sample of 33 total participants would be required to detect an effect size of .45, and a sample of 54 total participants would be required to detect an effect size of .37. Smith and colleagues reported small relations ($rs = .10$ to .28) between disgust reactivity and attitudes about other issues of sexual purity (e.g., premarital sex, pornography, abortion). Given an alpha of .05 and power of .80, a sample of 779 total participants would be required to detect an effect size of .10 and a sample of 95 total participants would be required to detect an effect size of .28. A total of 57 participants were included in the present study, thirty of whom were randomly allotted to the control group. As such, the present sample size is only sufficiently powered for medium to large effects.

A majority of participants were Caucasian/white ($n = 51$), self-described Christian ($n = 40$), and heterosexual ($n = 53$); there were no significant differences between the two experimental groups on any of the aforementioned demographic variables. There was a significant difference in gender distribution between the two experimental groups, X^2 (1) $= 6.22$, $p = .01$, with relatively more female than male participants in the control group ($n = 21$ and 9, respectively) compared to the disgust odor group ($n = 10$ and 17, respectively). The average age of the sample was 32.37 ($SD = 14.15$). The control group ($M = 35.33$, $SD = 15.59$) was marginally older than the disgust odor group ($M = 29.07$, $SD = 11.79$), F (1, 55) $= 2.87$, $p = .10$.

Measures

Disgust ratings. As a manipulation check, participants were asked to select a number ranging from 0 'Not at all' to 8 'Very much so' to indicate the degree to which they currently felt "disgusted, nauseated, repulsed."

Socio-political opinion questions. The Wilson-Patterson short-response format [17], which asks participants if they agree, disagree, or are uncertain on a range of currently salient public policy issues, was used to measure socio-political attitudes. Responses were recoded so that 1 = agree, 2 = uncertain, and 3 = disagree. Participants responded to 16 items concerning social, economic, regulatory, and international public policy. Even though our predictions were only concerned with sex-related attitudes, we opted to include other Wilson-Patterson items to allow for a direct comparison between our findings and those published by Smith and colleagues [2]. Eight items were worded such that higher scores reflected more agreement with conservative attitudes and eight items were worded such that lower scores reflected less agreement with conservative attitudes (see Table 1). Participants also completed four 5-point items that were specifically concerned with same sex marriage. The first of the aforementioned items was scaled from 1 'strongly oppose' to 5 'strongly favor' and was the same item used by Smith and colleagues [2] while the latter three items, which were developed for the present study, were scaled from 1 'strongly agree' to 5 'strongly disagree' (Table 1).

Conservatism Liberal Scale. The Conservatism Liberal Scale [23] is a one-item measure presenting the following: "We hear a lot of talk these days about liberals and conservatives. Here is a 7-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative." Participants are then asked to place themselves on the scale from 1 (extremely conservative) to 7 (extremely liberal).

Three Domains of Disgust Scale. The TDDS [13] is a 21-item self-report scale that measures three distinct domains of disgust sensitivity (moral, sexual, and pathogen). Each factor is represented by seven items measured on a zero to 6-point scale, ranging from 'not at all disgusting' to 'extremely disgusting'. The TDDS has strong psychometric properties [24] and internal consistencies of the TDDS factors were moderate within the present sample, ($\alpha_{\text{Moral Disgust}} = 0.83$), ($\alpha_{\text{Sexual Disgust}} = 0.82$), and ($\alpha_{\text{Pathogen Disgust}} = 0.79$).

Table 1. Descriptive data and summary of ANOVAs split by experimental groups.

Dependent variables	Condition	Mean (SD)	With Covariates *F* (ηp^2)	Without Covariates *F* (ηp^2)
5-Pt. Gay Marriage Items				
The federal government should support a constitutional amendment banning gay marriages*	Control	1.57 (1.04)	7.80 (.13)[b]	4.36 (.07)[a]
	Odor	2.33 (1.69)		
I should be allowed to marry whomever I want to, even if it is a member of the same sex	Control	1.70 (1.26)	15.42 (.23)[c]	7.38 (.12)[b]
	Odor	2.78 (1.72)		
If a close friend or family member were gay, I would support their right to having a same-sex marriage.	Control	1.53 (1.01)	17.29 (.25)[c]	9.70 (.15)[b]
	Odor	2.67 (1.69)		
Same sex marriage should be legalized nationwide	Control	1.80 (1.38)	13.88 (.21)[c]	5.49 (.09)[a]
	Odor	2.74 (1.65)		
Wilson-Patterson Items				
Gay marriage	Control	1.37 (.67)	12.54 (.19)[c]	4.75 (.08)[a]
	Odor	1.81 (.88)		
Premarital sex	Control	1.33 (.66)	6.50 (.11)[a]	2.95 (.05)
	Odor	1.71 (.95)		
Pornography	Control	1.83 (.87)	6.94 (.12)[b]	1.48 (.03)
	Odor	2.11 (.85)		
Abortion rights	Control	1.50 (.86)	1.47 (.03)	0.31 (.01)
	Odor	1.63 (.88)		
School prayer*	Control	2.30 (.92)	2.68 (.05)	1.40 (.03)
	Odor	2.00 (1.0)		
Illegal immigration	Control	2.30 (.88)	0.24 (.01)	0.78 (.01)
	Odor	2.48 (.64)		
Death penalty	Control	1.93 (.91)	0.00 (.00)	0.19 (.00)
	Odor	2.04 (.90)		
Biblical truth	Control	2.20 (.81)	11.24 (.18)[c]	6.04 (.10)[a]
	Odor	1.67 (.83)		
Welfare spending	Control	1.67 (.84)	0.95 (.02)	0.11 (.00)
	Odor	1.74 (.81)		
Tax cuts*	Control	1.73 (.74)	0.26 (.01)	0.11 (.00)
	Odor	1.67 (.78)		
Gun control	Control	1.73 (.86)	0.62 (.01)	0.43 (.01)
	Odor	1.89 (.93)		
Military spending*	Control	2.03 (.85)	0.61 (.01)	0.10 (.00)
	Odor	1.96 (.85)		
Warrantless searches*	Control	2.77 (.57)	.054 (.01)	1.77 (.03)
	Odor	2.93 (.27)		
Small government*	Control	1.77 (.77)	0.06 (.00)	0.02 (.00)
	Odor	1.74 (.66)		
Foreign aid	Control	1.70 (.84)	2.01 (.04)	0.65 (.01)
	Odor	1.89 (.93)		
Free trade	Control	1.43 (.73)	0.10 (.00)	0.08 (.00)
	Odor	1.48 (.58)		

Note. All ANCOVA $df = 1, 52$ and ANOVA $df = 1, 56$. "*" denotes items for which agreement (i.e., lower score) are consistent with a more conservative stance. Covariates included for ANCOVA analyses were gender, age, and moral disgust propensity.
Superscript "[a]" denotes $p \leq .05$, superscript "[b]" denotes $p \leq .01$, superscript "[c]" denoted $p \leq .001$.
doi:10.1371/journal.pone.0095572.t001

Results

A series of analysis of variance (ANOVA) tested dispositional differences between experimental groups on political ideology (conservative-liberal) and moral, sexual, and pathogen sensitivities. Results revealed that the no odor control group endorsed significantly more moral disgust sensitivity ($M = 31.14$, $SD = 5.90$) than the disgust odor group, ($M = 27.30$, $SD = 8.23$), $F(1, 55) = 4.16$, $p = .05$. The two experimental groups did not differ in terms of political ideology, sexual disgust, or pathogen disgust, (all $ps > .10$). The TDDS was completed several minutes after the disgust induction and in a corridor that was separate from the disgust odor infused room. If the disgust induction did have a lingering effect on TDDS scores, then participants would have likely also scored higher on the pathogen disgust subscale, which was not the case. It is, therefore, unlikely that the between group difference in moral disgust was due to the disgust induction.

An analysis of variance (ANOVA) showed that participants in the disgust odor condition reported more subjective disgust during the manipulation check ($M = 3.85$, $SD = 2.03$) than participants in the control condition ($M = 0.37$, $SD = 1.03$), $F(1, 52) = 68.56$, $p < .01$, $d = 2.16$. This effect remained significant after adding gender, age, and moral disgust sensitivity as covariates (see Participants section and the previous paragraph), $F(1, 52) = 48.28$ $p < .01$.

To account for between group differences in demographic and dispositional variables, a series of ANCOVAs – with gender, age, and moral disgust sensitivity as covariates – were carried out with the 4 sexual purity Wilson-Patterson and four 5-point gay marriage items. False discovery rates were controlled for using adaptive alpha correction procedures outlined by Benjamini and Hoochberg [25]. As expected, attitudes about gay marriage were affected by the disgust induction, with participants in the disgust odor condition reporting less agreement with gay marriage (as measured by the one Wilson-Patterson item and three 5-point items) and less opposition (and more support) to a constitutional ban on gay marriage ($ps < .01$). Also consistent with our predictions and the findings of Smith et al. [2], we found that participants in the disgust odor condition disagreed significantly more with the Wilson-Patterson premarital sex ($p \leq .01$) item. Consistent with our own prediction but in contrast with the findings of Smith and colleagues [2], participants in the disgust odor condition reported more disagreement with pornography ($p \leq .01$). While participants in the disgust odor condition did report less agreement with abortion rights, this effect was not significant ($p = .23$). Exploratory analyses revealed that participants in the disgust odor condition were in significantly more agreement with Biblical truth ($p < .001$). The experimental groups did not significantly differ on any of the remaining Wilson-Patterson items (all $ps > .05$). See Table 1 for a summary.

Group differences on all items were also analyzed with a series of ANOVAs. Consistent with ANCOVA results, univariate ANOVAs showed that participants in the disgust odor condition reported significantly less support for allowing individuals to marry whomever they wish ($p = .009$) and for allowing friends or family members to pursue same sex marriages ($p = .003$). Participants in the disgust odor condition also reported less agreement with the Wilson-Patterson gay marriage item ($p = .034$), less agreement with nationwide legalization of same-sex marriage ($p = .023$), and less opposition to an amendment banning gay marriage ($p = .041$); however, these effects were not significant after alpha correction. While participants in the disgust odor condition reported less agreement with premarital sex, pornography, and abortion, these effects were not statistically significant (all $ps > .05$). Consistent with ANCOVA results, univariate analyses showed that participants in the disgust odor condition reported more support for the Biblical truth item ($p = .017$). Scores on the remaining Wilson-Patterson items did not significantly differ between the two experimental groups (all $ps > .10$). See Table 1 for details.

Given that several effects were enhanced by the inclusion of covariates, correlations between age, moral disgust, and items measuring attitudes about sexual purity were inspected. There were strong correlations between age and several other variables, with older participants reporting more moral disgust sensitivity ($r = .43$, $p \leq .01$), less agreement with legalizing gay marriage ($r = .35$, $p \leq .01$), and less agreement with the Wilson-Patterson gay marriage item ($r = .34$, $p \leq .05$). Similarly, participants who reported more moral disgust sensitivity also reported more conservative attitudes for all items related to gay marriage ($rs \leq .30$ and $ps \leq .05$). A series of ANOVAs also revealed that female participants disagreed more with pornography compared with male participants. A two-way ANOVA with gender and experimental condition as the IVs and attitudes about pornography as the DV failed to show a significant gender by condition interaction $F(1, 53) = 0.83$, $p = .37$. None of the covariates were significantly related to attitudes about premarital sex ($ps > .05$). Taken as a whole, inspection of the relations between covariates and the dependent variables suggests that gender only affected attitudes about pornography, age and moral disgust sensitivity affected attitudes related to gay marriage, and no covariates were significantly related to attitudes about premarital sex.

Visual inspection of the distribution of the friend of family member same-sex marriage item – which demonstrated the largest effect size ($d = .82$) – suggested that the disgust odor induction caused attitudes to literally shift to the right (Figure 1). A smaller percentage of participants in the disgust odor condition strongly agreed with allowing friends or family members to pursue same-sex marriage and a larger percentage in this group strongly disagreed with such an idea. For example, zero participants in the no odor control group compared to 25.9% of participants in the disgust odor group strongly disagreed with allowing friends or family to pursue same-sex marriage. Interestingly, the experimental manipulation did not appear to dramatically affect the percentage of participants who endorsed more moderate attitudes (i.e., somewhat oppose to somewhat favor).

Discussion

Disgust has an important, causal relationship with political attitudes concerning sexual conduct. In an experiment manipulating odor-based disgust, participants exposed to the smell of butyric acid reported increased subjective disgust and more politically conservative attitudes concerning gay marriage, premarital sex, pornography, and Biblical truth. Rejection of gay marriage was a particularly strong response to the disgust-inducing odor, perhaps because of the connection between homosexuality and perceptions of sexual impurity [4,26]. The finding that belief in Biblical truth was greater among participants in the disgust odor condition was unexpected but is nonetheless consistent with previous work showing a relation between disgust and scrupulosity or being careful to avoid doing wrong. Indeed, Olatunji presented data that suggested a causal chain between disgust, scrupulosity, and conservative beliefs about homosexuality [10]. Future research should continue testing the causal effects of disgust on religious beliefs and the effects of religion on sexual attitudes.

We report non-significant findings for a variety of non-sexual socio-political attitudes (e.g., tax cuts, illegal immigration, etc.), but we do not interpret these null findings as evidencing an absence of relations between disgust and non-sexual socio-political attitudes.

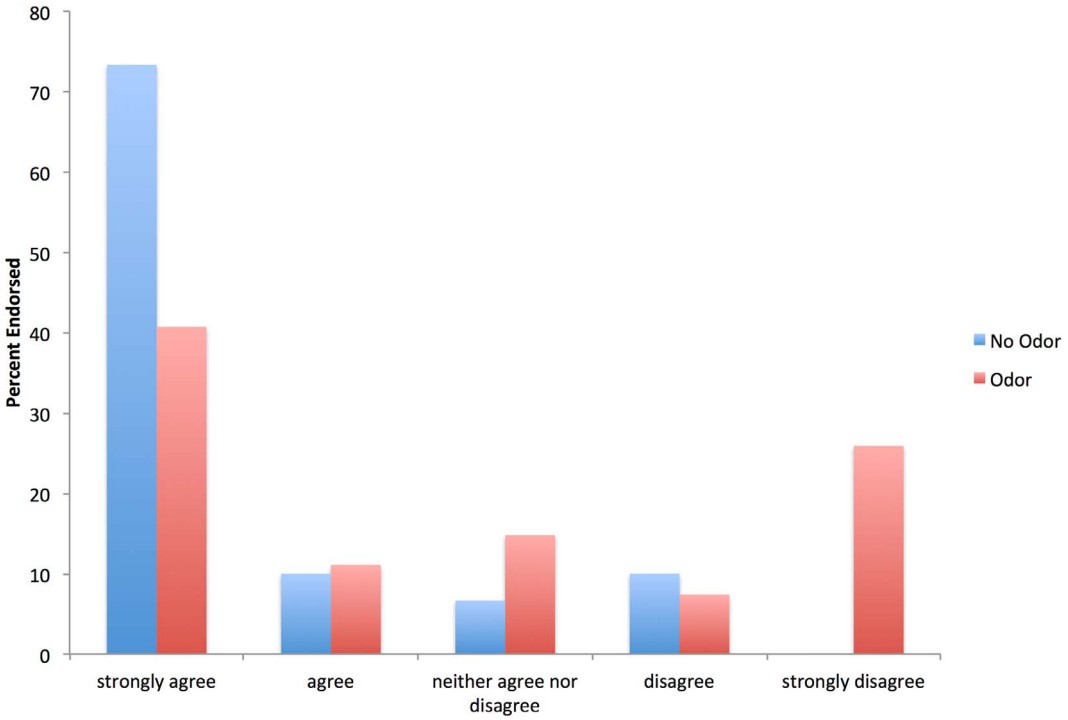


Figure 1. Attitudes about the following statement: "If a close friend of family member were gay, I would support their right to having a same-sex marriage."
doi:10.1371/journal.pone.0095572.g001

Replication with larger samples is needed before unconditional conclusions can be drawn, particularly with regard to smaller effects, which the present study was not appropriately powered to detect [e.g., abortion (see Participants section)]. Nonetheless, the present findings suggest that odor-induced disgust had a stronger effect on sexual attitudes than non-sexual attitudes (save for Biblical truth). The effects of disgust on sex-related attitudes, particularly those concerning gay marriage, are especially convincing when the present results are considered in the context of previous research [2,4,9]. For example, Smith and colleagues [2] reported very similar findings: greater changes in skin conductance following disgust provocation predicted stronger conservative views on gay marriage and premarital sex. However, the experimental approach of our study advances causal claims, particularly with regard to the effects of state disgust on views related to individual liberties pertaining to sexual orientation.

Participants in the no odor control condition were older, endorsed more moral disgust sensitivity, and were comprised of more females when compared to participants in the disgust odor condition. Gender was related to decreased agreement with pornography while age and moral disgust sensitivity were related to decreased agreement and support for same-sex marriage. It is, therefore, understandable that the inclusion of age, gender, and moral disgust sensitivity as covariates enhanced the magnitude of the experimental effect on attitudes about pornography and gay marriage. These results also suggest that extra emphasis should be placed on most of the ANCOVA results relative to the univariate analyses. Given the absence of relations between any of the covariates and premarital sex, it is difficult to determine precisely why the addition of covariates boosted this specific effect. The simplest explanation is that the addition of covariates reduced residual variance enough to push up effect size estimates and push down alpha estimates below significance thresholds. Regardless, the present findings, particularly those that were only significant with covariates (i.e., attitudes about premarital sex and pornography) must be interpreted with caution, as covariate analyses do no completely resolve sampling issues [27]. Future research should attempt to replicate the present findings with experimental groups that do not differ on important demographic and individual difference variables.

These data are consistent with theory delineated by Tybur and colleagues [13], which argues that disgust functions to decrease the occurrence (both in the self and society) of sexual behaviors that are perceived as increasing risk of pathogen transmission. Relatedly, and as noted by Haidt and Graham [11], conservative attitudes are driven not only by harm avoidance, but also by concerns about purity. According to these theories, shifts toward politically conservative views on sex may be basic, adaptive, and self-protective responses against perceived spread of pathogens or moral threats. When disgust is evoked, the behavioral immune system engages avoidance to prevent infection (e.g., less interpersonal contact [28]) and appears to moralize sexual conduct in ways that underlie conservative values of purity and sanctity [11,26]. As seen in the results of our study, it is possible that exposure to a disgusting odorant caused increased feelings of disgust, which in turn activated the harm avoidance system and motivated a desire for purity (cleanliness). Once these two systems were activated, it is possible that participants began to adopt attitudes that they perceived as decreasing social harm and/or increasing moral purity.

There is a growing literature indicating that disgust has important consequences for political views and policy preferences. In the research presented here, exposure to a disgusting odor caused greater endorsement of conservative views, including: rejecting gay marriage, restricting sex to marriage, disapproving of the use of pornography, and increased beliefs in Biblical truth.

Odor induced conservative shifts concerning gay marriage were particularly robust. It is possible that some forms of political conservatism, particularly those related to sex and sexuality, are basic and inherent in some populations and can readily emerge under threatening or taxing conditions [4,29,30].

Acknowledgments

We would like to thank Dr. Kevin Smith and Dr. Jeffry Lohr for their guidance and assistance with the preparation and publication of this manuscript.

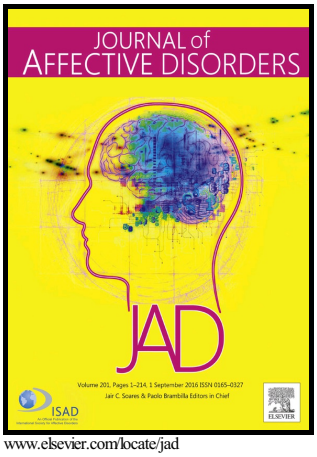
Author Contributions

Conceived and designed the experiments: TA PS. Performed the experiments: TA PS. Analyzed the data: TA. Contributed reagents/materials/analysis tools: TA PS. Wrote the paper: TA PS JB.

References

1. Hatemi P, McDermott R (2012) Policing the perimeter: Disgust and purity in democratic debate. PS: Political Science & Politics 45: 675–687.
2. Smith KB, Oxley D, Hibbing MV, Alford JR, Hibbing JR (2011) Disgust sensitivity and the neurophysiology of left-right political orientations. PLoS One 6: e25552.
3. Inbar Y, Pizarro D, Iyer R, Haidt J (2012) Disgust sensitivity, political conservatism, and voting. Social Psychological and Personality Science 3: 537–544.
4. Helzer EG, Pizarro DA (2011) Dirty liberals! reminders of physical cleanliness influence moral and political attitudes. Psychological science 22: 517–522.
5. Tybur J, Merriman L, Caldwell Hooper A, McDonald M, Navarrete C (2010) Extending the behavioral immune system to political psychology: Are political conservatism and disgust sensitivity really related. Evolutionary Psychology 8: 599–616.
6. Terrizzi JA, Shook NJ, Ventis WL (2010) Disgust: A predictor of social conservatism and prejudicial attitudes toward homosexuals. Personality and Individual Differences 49: 587–592.
7. Terrizzi Jr JA, Shook NJ, McDaniel MA (2012) The behavioral immune system and social conservatism: A meta-analysis. Evolution and Human Behavior 34: 99–108.
8. Balzer A, Jacobs CM (2011) Gender and physiological effects in connecting disgust to political preferences. Social Science Quarterly 92: 1297–1313.
9. Inbar Y, Pizarro DA, Bloom P (2009) Conservatives are more easily disgusted than liberals. Cognition & Emotion 23: 714–725.
10. Olatunji BO (2008) Disgust, scrupulosity and conservative attitudes about sex: Evidence for a mediational model of homophobia. Journal of Research in Personality 42: 1364–1369.
11. Haidt J, Graham J (2007) When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. Social Justice Research 20: 98–116.
12. Inbar Y, Pizarro DA, Knobe J, Bloom P (2009) Disgust sensitivity predicts intuitive disapproval of gays. Emotion 9: 435–439.
13. Tybur JM, Lieberman D, Griskevicius V (2009) Microbes, mating, and morality: Individual differences in three functional domains of disgust. J Pers Soc Psychol 97: 103–122.
14. Oaten M, Stevenson RJ, Case TI (2009) Disgust as a disease-avoidance mechanism. Psychol Bull 135: 303–321.
15. Toronchuk JA, Ellis GFR (2007) Disgust: Sensory affect or primary emotional system? Cognition and emotion 21: 1799–1818.
16. Schaller M, Murray DR (2011) Infectious disease and the creation of culture. Advances in culture and psychology 1: 99–151.
17. Wilson GD, Patterson JR (1968) A new measure of conservatism*. Br J Soc Clin Psychol 7: 264–269.
18. Koukounas E, McCabe MP (2001) Sexual and emotional variables influencing sexual response to erotica: A psychophysiological investigation. Arch Sex Behav 30: 393–408.
19. Rozin P, Lowery L, Ebert R (1994) Varieties of disgust faces and the structure of disgust. Journal of Personality and Social Psychology; Journal of Personality and Social Psychology 66: 870–881.
20. Kraut RE (1982) Social presence, facial feedback, and emotion. J Pers Soc Psychol 42: 853–863.
21. Soussignan R, Schaal B, Marlier L, Jiang T (1997) Facial and autonomic responses to biological and artificial olfactory stimuli in human neonates: Re-examining early hedonic discrimination of odors. Physiol Behav 62: 745–758.
22. Faul F, Erdfelder E, Lang A, Buchner A (2007) G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods 39: 175–191.
23. Knight K (1999) Liberalism and conservatism. In: Robinson JP, Shaver PR, Wrightsman LS, editors. Measures of political attitudes. Measures of social psychological attitudes, Vol. 2. San Diego, CA: Academic Press. pp. 59–158.
24. Olatunji BO, Adams T, Ciesielski B, David B, Sarawgi S, et al. (2012) The three domains of disgust scale factor structure, psychometric properties, and conceptual limitations. Assessment 19: 205–225.
25. Benjamini Y, Hochberg Y (2000) On the adaptive control of the false discovery rate in multiple testing with independent statistics. Journal of Educational and Behavioral Statistics 25: 60–83.
26. Graham J, Haidt J, Nosek BA (2009) Liberals and conservatives rely on different sets of moral foundations. J Pers Soc Psychol 96: 1029.
27. Miller GA, Chapman JP (2001) Misunderstanding analysis of covariance. J Abnorm Psychol 110: 40–48.
28. Schaller M, Park JH (2011) The behavioral immune system (and why it matters). Current Directions in Psychological Science 20: 99–103.
29. Eidelman S, Crandall CS, Goodman JA, Blanchar JC (2012) Low-effort thought promotes political conservatism. Person Soc Psychol Bull 38: 808–820.
30. Jost JT, Glaser J, Kruglanski AW, Sulloway FJ (2003) Political conservatism as motivated social cognition. Psychol Bull 129: 339–375.

Author's Accepted Manuscript



www.elsevier.com/locate/jad

Effects of behavioral activation on default mode network connectivity in subthreshold depression: a preliminary resting-state fMRI study

Satoshi Yokoyama, Yasumasa Okamoto, Koki Takagaki, Go Okada, Masahiro Takamura, Asako Mori, Syouichi Shiota, Naho Ichikawa, Ran Jinnin, Shigeto Yamawaki

PII: S0165-0327(17)30949-7
DOI: https://doi.org/10.1016/j.jad.2017.10.021
Reference: JAD9286

To appear in: *Journal of Affective Disorders*

Received date: 11 May 2017
Revised date: 26 September 2017
Accepted date: 4 October 2017

Cite this article as: Satoshi Yokoyama, Yasumasa Okamoto, Koki Takagaki, Go Okada, Masahiro Takamura, Asako Mori, Syouichi Shiota, Naho Ichikawa, Ran Jinnin and Shigeto Yamawaki, Effects of behavioral activation on default mode network connectivity in subthreshold depression: a preliminary resting-state fMRI study, *Journal of Affective Disorders,* https://doi.org/10.1016/j.jad.2017.10.021

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Effects of behavioral activation on default mode network connectivity in subthreshold depression: a preliminary resting-state fMRI study

Satoshi YOKOYAMA [1,2], Yasumasa OKAMOTO [1], Koki TAKAGAKI [1], Go OKADA[1], Masahiro TAKAMURA [1], Asako MORI [1], Syouichi SHIOTA [1,2], Naho ICHIKAWA [1], Ran JINNIN [1], Shigeto YAMAWAKI [1]

1) Department of Psychiatry and Neurosciences, Hiroshima University, Japan.

2) Research Fellow of Japan Society for the Promotion of Science, Japan.

Corresponding author:

Yasumasa Okamoto, MD, PhD.

Department of Psychiatry and Neurosciences, Graduate School of Biomedical and Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8551, Japan.

Email: oy@hiroshima-u.ac.jp

Accepted manuscript

Abstract:

Background: Subthreshold depression is a risk factor for major depressive disorder, and it is known to have a negative impact on quality of life (QOL). Although behavioral activation, which is one type of cognitive behavioral therapy, is an effective psychological intervention for subthreshold depression, neural mechanisms of behavioral activation are unclear. Enhanced functional connectivity between default mode network (DMN) and the other regions has been

ACCEPTED MANUSCRIPT

demonstrated in participants with subthreshold depression. The purpose of this study was to examine the effects of behavioral activation on DMN abnormalities by using resting-state functional MRI (rs-fMRI).

Methods: Participants with subthreshold depression (N=40) were randomly assigned to either an intervention group or a non-intervention group. They were scanned using rs-fMRI before and after the intervention. Independent component analysis indicated three subnetworks of the DMN.

Results: Analyzing intervention effects on functional connectivity of each subnetwork indicated that connectivity of the anterior DMN subnetwork with the dorsal anterior cingulate was reduced after the intervention. Moreover, this reduction was correlated with an increase in health-related QOL.

Limitations: We did not compare the findings with healthy participants. Further research should be conducted by including healthy controls to verify the results of this study.

Conclusions: Mechanisms of behavioral activation might be related to enhanced ability to independently use the dACC and the DMN, which increases an attention control to positive external stimuli. This is the first study to investigate neural mechanisms of behavioral activation using rs-fMRI.

Accepted manuscript

Key words: Subthreshold depression; Behavioral activation; Resting-state fMRI; Default mode network; Quality of life

1. Introduction

ACCEPTED MANUSCRIPT

Recently, it has been suggested that subthreshold depression, which is defined as clinically significant depressive symptoms not meeting diagnostic criteria for major depressive disorder (MDD: Pincus et al., 1999), is a potential risk factor for MDD (Bertha and Balázs, 2013; Jinnin et al., 2017). This condition is frequently observed among adolescents, and it is known to have a negative impact on quality of life (QOL: Bertha and Balázs, 2013). Although symptoms of subthreshold depression are less severe than those of MDD, the resulting impairments of daily life functions, including the deficit in QOL is similar to that in individuals with MDD, and significantly worse than in healthy people (Nierenberg et al., 2010). As a result, it is considered essential to elucidate the pathogenic mechanisms of subthreshold depression and to develop effective interventions. In particular, psychotherapy might be preferable for certain groups such as subthreshold cases or adolescents.

A meta-analysis (Cuijpers et al., 2007) suggested that cognitive behavioral therapy (CBT) is an effective intervention for subthreshold depression. Moreover, our previous study (Takagaki et al., 2014) on behavioral characteristics of subthreshold depression demonstrated that people with subthreshold depression are hyposensitive to reward. Therefore, a primary goal of interventions for subthreshold depression is to increase access to positively reinforcing events and activities. Behavioral activation is a CBT technique used for increasing the frequency of positive activities that lead to reinforcements in their life (Jacobson et al., 2001). The efficacy of behavioral activation matched to behavioral characteristics of clients with subthreshold depression has been confirmed in our recent randomized controlled trial (Takagaki et al., 2016), which demonstrated that behavioral activation was effective for improving depressive symptoms as well as clinical variables such as health-related QOL and the exposure to environmental

Accepted manuscript

ACCEPTED MANUSCRIPT

rewards (Takagaki et al., 2016). However, the underlying neural mechanisms of behavioral activation that are associated with the improvement of clinical variables are not fully understood.

A number of resting-state functional magnetic resonance imaging (rs-fMRI) studies that examined functional abnormalities of the brain and therapeutic responses in MDD have reported that depressed patients exhibit hyperactivity of the default mode network (DMN), which is composed of the medial prefrontal cortex (MPFC), the precuneus/posterior cingulate cortex (PCC), the inferior parietal lobule (IPL) and the hippocampal formation (Buckner et al., 2008). According to meta-analysis (Sundermann et al., 2014b), increased functional connectivity in the DMN regions, including the region between the anterior cingulate cortex (ACC) and the medial frontal cortex, as well as the region between the precuneus and PCC, was associated with depression. It is considered that the DMN is associated with self-referential processing (Lemogne et al., 2010). Patients with MDD often show abnormal self-focus biases, such as depressive ruminations (Grimm et al., 2009) and they focus on negative self-relevant information (Nolen-Hoeksema, 2000). Moreover, the functional connectivity between the DMN, including the ventrolateral and ventromedial prefrontal cortices, superior temporal gyri and limbic areas, and subgenual ACC, has the highest power to discriminate MDD patients from healthy controls (Zeng et al., 2014). Therefore, the DMN connectivity is considered to play a central role in the physiopathology of depression (Li et al., 2013).

Moreover, DMN activity of depressive people is known to be abnormally linked to other brain regions. Hwang et al. (2016) demonstrated that subthreshold depression was associated with enhanced functional connectivity between the DMN and regions outside the DMN, similar

Accepted manuscript

ACCEPTED MANUSCRIPT

to patients with MDD (Greicius et al., 2007). Moreover, changes in functional connectivity of the DMN were observed in an early stages of developing MDD, which might be associated with the behavioral characteristics of subthreshold depression. Therefore, investigating whether such abnormalities in the DMN are changed by behavior activation is important for understanding the neural mechanisms of behavior activation in subthreshold depression. However, to date, there are no published studies that have specifically focused on changes in spontaneous brain functions during the course of psychological treatment of subthreshold depression.

The purpose of this study was to examine the neural effects of behavioral activation on the DMN in people with subthreshold depression using rs-fMRI. The DMN can be divided along the anterior-posterior and inferior-superior axes (Allen et al., 2011). Li et al. (2013) investigated changes in the DMN subnetworks of MDD patients following antidepressant treatment, by using independent component analysis (ICA) approach. They reported that abnormalities in the posterior subnetwork were normalized after treatment, whereas there were no changes in the anterior subnetwork. CBT techniques including behavioral activation are mainly known to affect the frontal areas of the brain including the dorsal, ventral, and medial frontal cortices (DeRubeis et al., 2008; Goldapple et al., 2004; Ritchey et al., 2011; Yoshimura et al., 2014). A recent study (Shou et al., 2017) and a review (Mason et al., 2016) of predictors and mechanisms of CBT have also reported that evidence of functional connectivity resulting from CBT was most commonly found in fronto-limbic circuitry. Therefore, we hypothesize that successful behavioral activation would induce changes in functional connectivity of anterior DMN subnetwork. Based on this hypothesis and the method of previous studies (Li et al., 2013), we used only the anterior

Accepted manuscript

ACCEPTED MANUSCRIPT

and posterior subnetworks in the analyses. We did not focus on the temporal lobes, which are also considered to be key regions within the DMN.

2. Methods

2.1. Participants

Participants with subthreshold depression were recruited among freshmen attending Hiroshima University (age = 18 to 19) as diagnosed by the Beck's Depression Inventory-II (BDI-II: Kojima and Furukawa, 2003) and the Japanese version of the Composite International Diagnostic Interview (CIDI: Kawakami et al., 2005). Participants scoring 10 or more points on the BDI-II at all three time points [screening (week 0): $M \pm SD$ = 19.30±4.27, range=14 to 31, conducing the CIDI (week 20): $M \pm SD$ =16.28±4.53, range=10 to 30, and pre-intervention fMRI scanning (week 25): $M \pm SD$ =15.4±3.96, range= 10 to 27] were considered as having subthreshold depression. Participants that met the following criteria after the CIDI conducted by evaluators blind to the subject's assigned group were excluded from this study: (a) major depressive episode within the past one year, (b) a lifetime history of bipolar disorders, (c) taking psychopharmacological or psychological treatment within the past year, and (e) possibility of acute suicide attempts. The remaining participants with subthreshold depression (n =40) were included in the study and were randomly assigned to either an intervention (n =19) or a non-intervention group (n =21).

The intervention group took part in five, weekly, behavioral activation sessions that were conducted by a trained therapist for 60 min per session. The Behavioral activation program (Takagaki et al., 2016) was focused on expanding positively reinforcing activities in order to

Accepted manuscript

ACCEPTED MANUSCRIPT

increase the rate of response-contingent positive reinforcement. First, participants learned to assess their long and short-term goals, and to develop activity monitoring based on psycho-education about depression and behavioral activation (Session 1). Next, they constructed a behavioral ranking consisting of 10 tasks that gave participants a sense of reward (Session 2). Sessions 2–4 focused on behavioral experiments and increasing scheduled activities. In the last session (Session 5), they reviewed behavioral activation and developed a coping behavior plan for stressful situations in the future. Sessions 1-4 included homework. We have already conducted a randomized controlled trial to confirm the effectiveness of this program for late adolescents with subthreshold depression (Takagaki et al., 2016). This program is highly effective in significantly reducing depressive symptoms (Hedges' g =-0.90) and increasing health-related QOL (Hedges' g =0.57-0.77). Participants in the intervention group completed clinical measures, and rs-fMRI scanning was conducted at pre and post intervention. None of the subjects were absent for any of the sessions. They completed all of their homework assignments. Furthermore, a trained therapist conducted the intervention, and adherence to the protocol was assessed as being 100% by two supervisors (for details see, Takagaki et al., 2016). The non-intervention group also participated identically in the study, with the exception of the intervention program. All participants provided their written informed consent before participating in the study. The study was approved by the ethics committee of Hiroshima University.

Accepted manuscript

2.2. Clinical measures

2.2.1. Japanese version of the Beck's Depression Inventory-II (BDI-II)

ACCEPTED MANUSCRIPT

The BDI-II is widely used scale for assessing symptoms that consist of 21 self-report items. The items are scored using a 4-point scale. The Japanese version of the scale has demonstrated reliability and validity (Kojima and Furukawa, 2003).

2.2.2. Japanese version of the Behavioral Activation for Depression Scale (BADS)

The BADS (Kanter et al., 2007) measures the frequency of activation, and consists of 25 items that are rated on a 7-point scale (0: Not at all to 6: Completely). Four subscales of BADS are Activation (AC: 7 items) representing focused, goal-directed activation and completion of scheduled activities, Avoidance/Rumination (AR: 8 items) representing avoidance of negative aversive states and engaging in rumination, Work/School Impairment (WS: 5 items), and Social Impairment (SI: 5 items). We used only the BADS-AC subscale, because our program focused on increasing access to positively reinforcing activities to improve the rate of response-contingent positive reinforcement. The Japanese version of the BADS has demonstrated reliability and validity (Takagaki et al., 2013).

Accepted manuscript

2.2.3. Japanese version of Environmental Reward Observation Scale (EROS)

The EROS (Armento and Hopko, 2007) is used to assess exposure to environmental rewards deemed essential for increasing response-contingent positive reinforcement. Items were composed of behaviors and positive affects as a consequence of rewarding environmental experiences (Armento and Hopko, 2007). The original version of the scale consists of 10 items that are scored on a 4-point scale (1: Strongly disagree to 4: Strongly agree).The Japanese version of EROS has demonstrated reliability and validity (Kunisato et al., 2011).

ACCEPTED MANUSCRIPT

2.2.4. Japanese EuroQOL-5 Dimension (EQ-5D)

The EQ-5D (Euroqol Group, 1990) was developed to assess health-related quality of life and consists of 5 self-report items (3-point scale) and a visual analogue scale (VAS). The self-report items reflect health scored from 0 (death) to 1 (perfect health). VAS evaluates health status as ranging from 0 (worst imaginable health state) to 100 (best imaginable health). The Japanese version of the EQ-5D has demonstrated reliability and validity (Nishimura et al., 1998).

2.3. fMRI data acquisition

Imaging data were acquired using a 3.0 T SIEMENS MAGNETOM with a 12-channel head coil (Siemens, Munich, Germany). Resting-state echo planar imaging (EPI) acquired for approximately 10 minutes for each participant (40 slices; TR/TE = 2500/30 ms; thickness = 3.2 mm; FOV = 212 x 212 mm; flip angle = 80°; matrix = 64 x 64; 244 volumes). During the rs-fMRI scans, participants were instructed to keep looking at a central fixation point, to keep still, to stay awake, and not to think about anything specific. Anatomic T1-weighted images were acquired using an MPRAGE sequence (192 slices; TR/TE = 2300/2.98 ms; thickness = 1 mm; FOV = 256 x 256 x 192 mm; flip angle = 9°; matrix = 256 x 256). All participants had not used nicotine or alcohol because the law prohibits people under 20 from taking these substances in Japan, and these compliances were confirmed by the investigators. Moreover, participants had not used caffeine at least 3 days prior to the scanning session.

Accepted manuscript

ACCEPTED MANUSCRIPT

2.4 fMRI data preprocessing

Preprocessing was performed via Statistical Parametric Mapping software (SPM8: www.fil.ion.ucl.ac.uk/spm). The first four volumes of rs-fMRI data were discarded to ensure signal equilibrium and participants' adaptation to scanning noise. After this the remaining 240 images were first time-corrected for each participants (slice timing) and subsequently realigned by using rigid body transformation. The T1-weighted anatomical volume was coregistered to the mean image created by the realignment procedure and was normalized to the Montreal Neurological Institute (MNI) space using a tri-linear interpolation. All functional images were then smoothed using 6-mm full-width at half-maximum Gaussian kernel.

2.5 Independent component analysis and selection on networks-of-interest

We performed group-ICA using the GIFT-software (http://icatb.sourceforge.net/l). ICA is a data-driven technique for identifying functional connectivity patterns in resting-state fMRI (Beckmann, 2012). As head motion may have both noise and neuronal effects on functional connectivity measures (Satterthwaite et al., 2012; Ling-Li. Zeng et al., 2014), we confirmed that there were no differences in head motion parameters by comparing the groups as well as pre-intervention and post-intervention before performing ICA. Based on previous studies (Allen et al., 2011; Manoliu et al., 2013), we decomposed our data into 75 components based on the infomax-algorithm (Calhoun et al., 2001). It has been demonstrated that decomposing into 75 components is optimal for detecting between-group differences and for avoiding false positive results (Abou-Elseoud et al., 2010). The reliability and robustness of the estimated components were subsequently maximized by running 20 iterations of the ICA using the ICASSO toolbox

Accepted manuscript

ACCEPTED MANUSCRIPT

(Himberg et al., 2004). Then, to identify DMN components, we performed multiple spatial regression analyses on spatial maps of the 75 components using three DMN templates, reflecting anterior, inferior posterior, and superior posterior subnetworks. These templates were chosen from T-maps of 28 resting-state networks available on Internet (Allen et al., 2011: http://mialab.mrn.org/data/hcp/RSN_HC_unthresholded_tmaps.nii). We selected the three DMN components having the highest correlation coefficients (component 13: 0.47, $p < .05$; component 56: 0.45, $p < .05$; component 39: 0.32, $p < .05$) with the templates. The amplitude of each component is known to reflect the contribution of each voxel to the distribution and the coherent activity within that component and therefore, it is interpreted as a functional connectivity (Manoliu et al., 2013).

Accepted manuscript

2.6. Statistical analyses

Voxel-wise one-sample *t* tests were used to identify the spatial pattern of each DMN component (multiple comparisons with $p < .05$ family-wise-error correction on the peak- and cluster-level). We then used the mask obtained from one-sample *t* test results corresponding to each components for the following brain analysis. Each mask was created by combining the spatial maps of the two groups obtained from one-sample *t* test results for each components within intervention and non-Intervention groups, respectively. Intervention effects on clinical states such as depressive symptoms were examined by using two-way analyses of covariance (ANCOVAs), with groups and time points as independent variables and clinical scores (e.g. BDI-II) as dependent variables. Age and gender were included in the analyzed model as covariates. Bonferroni corrected post-hoc tests ($p < .05$) were conducted to specify significant

group and time effects. For the whole brain analysis, we decided on a cut off threshold of p <.001 uncorrected and a spatial extent of 10 voxels (k >=10) to avoid Type II errors (Lieberman and Cunningham, 2009). And similar ANCOVA models were applied to the amplitude of the voxels within each component in order to examine the effect of the intervention on each DMN component. Finally, we calculated correlations between the degree of changes in functional connectivity, by subtracting the pre-component image from the post-component image using ImCalc on SPM within the resulting areas of ANCOVAs interaction, and clinical scores within each group. The statistical threshold for these correlation analyses was set at $p < .05$.

ACCEPTED MANUSCRIPT

3. Results

3.1. Subnetworks of the DMN in subthreshold depression

Three components of the DMN subnetworks were identified in all participants (Figure 1): component 13 representing the anterior DMN (aDMN) had the highest amplitude in ACC and MPFC, component 56 representing the inferior posterior DMN (ipDMN) had the highest amplitude in PCC, and component 39 representing the superior posterior DMN (spDMN) had the highest amplitude in precuneus.

Accepted manuscript

3.2. Intervention effects on clinical scores

Results of ANCOVAs conducted on scores of most clinical scores assessment scales (BDI-II, BADS-AC EROS and VAS of EQ-5D) showed significant interactional effects between group and time (Table 1; BDI-II: $F(1, 36) = 7.34$, $p < .05$, *partial* $\eta^2 = 0.17$; BADS-AC: $F(1, 36) = 26.00$, $p < .001$, *partial* $\eta^2 = 0.42$; EROS: $F(1, 36) = 6.67$, $p < .05$, *partial* $\eta^2 = 0.16$; VAS of EQ-5D:

ACCEPTED MANUSCRIPT

$F(1, 36) = 12.23$, $p < .001$, *partial* $\eta^2 = 0.25$). Post-hoc analysis with Bonferroni correction indicated significantly lower BDI-II and higher EROS and VAS scores at post-intervention in the intervention group compared to the non-intervention group (BDI-II: $p < .01$, *partial* $\eta^2 = 0.14$; EORS: $p < .05$, *partial* $\eta^2 = 0.12$; VAS of EQ-5D: $p < .001$, *partial* $\eta^2 = 0.30$). There were also significant improvements in the intervention group at post-intervention compared to pre-intervention (BDI-II: $p < .001$, *partial* $\eta^2 = 0.36$; BADS-AC: $p < .001$, *partial* $\eta^2 = 0.55$; EROS: $p < .01$, *partial* $\eta^2 = 0.24$; VAS of EQ-5D: $p < .01$, *partial* $\eta^2 = 0.19$). BADS-AC score of the non-intervention group at pre-intervention was higher than that of the intervention group ($p < .01$, *partial* $\eta^2 = 0.18$). Therefore, we used the BADS-AC score at pre-intervention as a covariate in the subsequent analysis. These ANCOVAs are described in the next section. Moreover, there was a significant reduction of VAS in the non-intervention group ($p < .05$, *partial* $\eta^2 = 0.11$).

3.3. Changes in DMN subnetworks after behavioral activation

The results of ANCOVAs are shown in Figure 2. It can be seen that there is a significant interactional effect in the dorsal ACC (dACC) within the aDMN component ($p < .001$ uncorrected, $k >= 10$). Figure 2B shows the coherent activity of the dACC within the aDMN component, which implies a functional connectivity between the dACC and this component, in both groups at each time point. Post-hoc analysis revealed that the aDMN-dACC connectivity was significantly reduced in the intervention group ($p < .001$ uncorrected, $k >= 10$). There were no significant interactions, or main effects of group or time in the ipDMN or the spDMN.

3.4. Correlations of the DMN subnetworks with clinical scores

Accepted manuscript

ACCEPTED MANUSCRIPT

We conducted correlation analyses with differences between pre and post intervention in aDMN-dACC functional connectivity and clinical data in order to explore associations between changes in aDMN-dACC connectivity and treatment response. We found a significant negative correlations between change in aDMN-dACC functional connectivity and VAS changes of EQ-5D in the intervention group (Figure 3: $r = -.61$, $p < .01$), which remained significant after controlling for BDI-II score ($r = -.62$, $p < .01$). Other clinical scores were not correlated with change of the aDMN-dACC functional connectivity either both group (BDI-II: $r = .12$, *n.s.*, $r = -.17$, *n.s.*; BADS-AC: $r = -.02$, *n.s.*, $r = .36$, *n.s.*; EROS: $r = .08$, *n.s.*, $r = .06$, *n.s.*; EQ-5D: $r = -.23$, *n.s.*, $r = .10$, *n.s.*; intervention and control group, respectively).

Accepted manuscript

4. Discussions

There were two main findings of this study. Firstly, among the three subnetworks of the DMN that identified by ICA in participants with subthreshold depression, behavioral activation might have specifically reduced the functional connectivity in the aDMN subnetwork with the dACC at least compared to non-intervention, although it is not possible to make any inferences about causation. Secondly, the reduction in the connectivity between the aDMN and the dACC was negatively correlated with the increase in subjective health-related QOL of the intervention group. This is the first study using rs-fMRI to report the effect of behavioral activation on spontaneous brain functions in people with subthreshold depression.

Decreased functional connectivity between the aDMN and the dACC following behavioral activation, implies that activity pattern of the dACC was de-synchronized with the time series of

ACCEPTED MANUSCRIPT

our aDMN component. On the other hand, behavioral activation had no effect on posterior subnetworks of the DMN (i.e. the ipDMN, or the spDMN). Antidepressants and psychotherapy (e.g. cognitive therapy) is known to work through different mechanisms (DeRubeis et al., 2008). A number of previous studies that examined neural responses to CBT have demonstrated functional brain changes in the dorsal, ventral, and medial frontal cortices (DeRubeis et al., 2008; Goldapple et al., 2004; Ritchey et al., 2011; Shou et al., 2017; Yoshimura et al., 2014) and dorsal/pregenual ACC (Fu et al., 2013; Konarski et al., 2009) related to treatment. CBT-related changes were most commonly observed in the DMN regions, which include the MPFC, PCC, hippocampus, and ACC sub-regions including dorsal, subgenual, and ventral (as review: Franklin et al., 2015). As already mentioned, DMN connectivity is deeply involved with depressive self-referential processing such as rumination (Li et al., 2013; Sundermann et al., 2014a; Ling-Li Zeng et al., 2014). Moreover, the decrease in dACC activity at resting-state following CBT could be conceptualized as the result of CBT improving a putative 'dorsal cognitive circuit', important in cognitive control and effortful regulation of emotion (Franklin et al., 2015). These findings might suggest that there is a neural circuit of behavioral activation in common with CBT. Whereas one previous study has shown that antidepressants act on the posterior DMN (Li et al., 2013). Goldapple et al. (2004) reported that depressed subjects have shown decreases of activation in the dorsal, medial and ventral prefrontal cortices, in addition to increases of activation in the hippocampus and in the dorsal cingulate cortex after CBT. They also obtained different patterns after paroxetine treatment, including increases in the prefrontal cortex and decreases in the hippocampus and subgenual cingulate. A subsequent study by Kennedy et al. (2007) compared CBT to venlafaxine and replicated these findings, which again suggested different

Accepted manuscript

types of neural changes following CBT and pharmacotherapy. Unlike other psychiatric disorders, including obsessive-compulsive and panic disorders, the mechanisms of pharmacological medication and psychotherapeutic treatment in depression seem to be divergent, or only partly overlapping (Barsaglini et al., 2014). Our results suggested that behavioral activation might act differently to pharmacotherapy, by changing the anterior subnetwork of the DMN, which supported our hypothesis and findings of previous studies investigating the neural responses to CBT.

Generically speaking, the dACC is known as a core region of the salience network (SN: Menon and Uddin, 2010), which is typically distinguished from the DMN. However, in this study, ICA analysis of our data on participants with subthreshold depression identified the DMN and the dACC as identical component. ICA is a data-driven analysis method and our data points were composed mainly of subjects with subthreshold depression who had not been subjected to an intervention (pre-intervention and non-intervention conditions), which resulted in the dACC being included in the aDMN "component". It has been reported that increased functional connectivity is observed between the DMN and other regions, which the included bilateral middle frontal gyrus, left superior frontal gyrus, right inferior frontal gyrus, ventral striatum, thalamus, parahippocampus, and insula, in people with subthreshold depression (Hwang et al., 2016) and that functional connectivity between the DMN and the SN is stronger in patients with MDD than in healthy participants (Manoliu et al., 2013). These findings suggested that enhanced functional connectivity between the DMN and the dACC in the SN might be closely associated with the pathology of subthreshold depression. The SN, which anchors the dACC and the anterior insula, is known to be involved in the detection of salient stimuli and events(Seeley et al., 2007) in order to guide behavior (the ACC is involved in action selection in particular:

ACCEPTED MANUSCRIPT
Accepted manuscript

ACCEPTED MANUSCRIPT

Rushworth, (2008)), and play a central role in biased negative informative processing in depression (Hamilton et al., 2012). High aDMN-dACC connectivity during the resting-state might reflect co-working of internal self-reference processing (DMN) and the detection of these internal stimuli (SN). In addition, the SN critically contributes to switching between the DMN and the executive network, so the dACC and anterior insula are uniquely positioned to initiate control signals that deactivate the DMN and activate the executive network (Seeley et al., 2007). Therefore, individuals with enhanced connectivity of the two networks, DMN and SN, might be contribute to directing attention to internal-self stimuli, without awareness of external stimuli as a consequence of deactivation in the executive network that contributes to the control of active attention. Again, our results revealed that the contribution of the dACC absorbed in the aDMN component has decreased following behavioral activation. This indicated that the dACC might not have been included in the aDMN component after intervention, as it is originally defined. Our behavioral activation program was developed for the purpose of increasing participants' access to positive activities, which was accompanied by reinforcements from their environment (Takagaki et al., 2016). Results indicated an increase in these activities and an increase in exposure to environmental rewards (Table 1), suggested that this purpose was adequately achieved. The dACC would be activated when detecting the reward and selecting behavior, and would be required to control aDMN deactivation. Such practices might have increased opportunities for activating the dACC and deactivating the DMN. The mechanisms of behavioral activation might be related to enhanced ability to independently use the dACC and the DMN, which advanced the control of attention favoring positive stimuli from the external

Accepted manuscript

ACCEPTED MANUSCRIPT

environment. The use of a placebo and comparison with other psychotherapies are expected to further clarify this interesting mechanism.

The dACC plays a critical role in the development of cognitive control skill that are required for adaptive functioning (Ordaz et al., 2013). Moreover, it is known that dACC activity increases with aging (Rubia, 2013) from childhood (9-17 years) to adulthood (18-26 years). One model of adolescent brain development (Crone and Dahl, 2012) describes the gradual development of the cognitive control system (dACC, dorsolateral PFC and the parietal cortex) interacting with social-affective processing and leading to increased adaptive adult functioning such as social motivation and tendencies to explore, take risks and try new things. Additionally, within-functional networks (including the DMN) are enhanced during development from early adolescence to late adolescence (Fair et al., 2008; Kelly et al., 2009; Saverino et al., 2015). A recent study on enhanced connectivity of depressed youth in remission (Jacobs et al., 2014) has also reported hyper-connectivity of the DMN and SN, whereas young adolescents with depression did not shown significant differences in resting-state DMN connectivity compared to health young adolescents (Pannekoek et al., 2014). Because the participants in this study were late adolescents, changes in DMN-dACC connectivity might be critical to many domains of adaptive adult functioning. Therefore, we suggest that future studies investigate other age groups such as children and adults with and without depression by using the research design developed in this study.

Accepted manuscript

Certain late adolescents with subthreshold depression show increased depressive symptoms and develop a major depressive episode during the natural course of one year (Jinnin et al., 2017). Therefore, it is significant that behavioral activation results in positive changes in many clinical variables for late adolescents. Results also indicated an association between

changes in connectivity within the aDMN component and VAS score of EQ-5D, which is a measure of health-related QOL. This was not the case between connectivity and other clinical variables such as the BDI-II, BADS, or EROS. Recently, QOL has been recognized as a more pertinent outcome measure because improvement of QOL is important in clinical settings as well as improvements of symptoms (Satomura et al., 2014). Individuals with subthreshold depression have lower depressive symptoms than patients with depression. Nevertheless, functional impairments in subthreshold depression are severe and serious (Bertha and Balázs, 2013; Nierenberg et al., 2010). Our previous study (Takagaki et al., 2016) reported that behavioral activation for subthreshold depression improves health-related QOL. Therefore, it is significant that changes in functional connectivity, possible caused by behavioral activation, were linked to the positive effects experienced by the participants. Faget-Agius et al. (2016) investigated brain functions associated with the QOL of schizophrenia patients using whole-brain single photon emission computed tomography in the resting-state. Their result indicated that not only the prefrontal area but also areas of the DMN such as precuneus (Buckner et al., 2008) and areas co-activating with the DMN such as parahippocampal gyrus (Greicius et al., 2003) were related to QOL. Although there are still few studies on the relationship between QOL and brain functions, our results showing that a change in the aDMN-dACC connectivity was associated with health-related QOL also where quite consistent with their results. As discussed in the previous paragraph, the reduction in aDMN-dACC functional connectivity would have advanced the control of attention to positive stimuli, and both the DMN and the dACC are associated with better prediction of treatment outcome (Fu et al., 2013; Ritchey et al., 2011). It is possible that increased awareness of positive stimuli in the

ACCEPTED MANUSCRIPT
Accepted manuscript

ACCEPTED MANUSCRIPT

reward acquisition process, which was associated with a reduction in aDMN-dACC connectivity following our behavioral activation program might result in an enhanced outlook on health in life.

The results of this study are constrained by several limitations. Firstly, we did not compare the enhanced functional connectivity between the DMN and the dACC found in people with subthreshold depression with healthy participants. Moreover, we used thresholds recommended for balancing Type I and Type II error (Lieberman and Cunningham, 2009). Our findings must be considered preliminary because we did not use correction methods for multiple comparisons. The detection of intervention effects should be re-examined using more severe thresholds or correction for multiple comparisons to avoid Type I errors. Although abnormal change in functional connectivity of the DMN in people with subthreshold depression have already demonstrated in a previous study (Hwang et al., 2016), further research should be conducted by including healthy controls to verify the results of this study. Second, we did not examine the function of aDMN-dACC using measures of directing attention to internal-self stimuli, or awareness of external stimuli. There is a need to conduct a mediation analysis, particularly with regard to the association with health-related QOL. Nevertheless, we could not conduct such a mediation analysis because of the small sample size. The assessments of such psychological functions are necessary in order to examine the consideration that we have stated in the discussion. Thirdly, the lack of a placebo group makes it difficult to attribute the findings of this study to any specific treatment effects. This is because the effects of other specific factors, known to be related to the treatment efficacy of psychotherapy independently of its orientation (CBT, Psychodynamic treatment etc.) cannot be discounted. Fourth, we focused on subthreshold

Accepted manuscript

depression and did not include clinical course data (e.g. development of MDD) in our analyses. The study how behavioral activation and brain function are related to future depression is also required in the future. Finally, our sample was not population-based, and participants were students of the same university, limiting the generalizability of our findings. Nevertheless, the homogeneity of the sample might also be a strength of this study because it diminished the possible effects of demographic variables as confounds.

ACCEPTED MANUSCRIPT

5. Conclusions

Behavioral activation induced a specific reduction in functional connectivity between the aDMN subnetwork and the dACC compared to the non-intervention group, which was negatively correlated with an increase of subjective health-related QOL in the intervention group. Mechanisms of the therapeutic efficacy of behavioral activation might be related to enhancing the ability for independently using the dACC and the DMN, which might advance an attention control directed at positive stimuli in the external environment. These findings contribute to our understanding of neural mechanisms of behavioral activation on subthreshold depression.

Accepted manuscript

Author Disclosure

Contributors

Author SY1 undertook the data analysis and wrote the manuscript. Authors SY1, KT, MT, AM, and SS acquired the data. Authors SY1, YO, KT, RJ and SY2 designed the study. Authors YO, MT, NI and GO supervised the data analysis. All authors contributed to and have approved the final manuscript.

Funding source

ACCEPTED MANUSCRIPT

This study conducted in at the Hiroshima were supported by a Grant-in-Aid for Scientific Research on Innovative Areas, Grant Number 16H06395 and 16H06399, from JSPS and Grant Number 23118004 from the Ministry of Education, Culture, Sports, Science and Technology, Japan. Part of this work was also conducted under the "Brain-machine interface development" under the Strategic Research Program for Brain Sciences (SRPBS) of the Ministry of Education, Culture, Sports, Science and Technology of Japan (grant no. 15dm0107050h0003), Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS) by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (grant no. 15dm0207012h0002), and Japan Society for Promotion of Science (JSPS) KAKENHI (grant no. 26461748).

Acknowledgements

The authors acknowledge the contributions of the TEXT Co. (ratnin@mac.com), Ltd, who proofread this article.

Accepted manuscript

Conflicts of Interest

Conflicts of interest: none

References

Abou-Elseoud, A., Starck, T., Remes, J., Nikkinen, J., Tervonen, O., Kiviniemi, V., 2010. The effect of model order selection in group PICA. Hum. Brain Mapp. 31, 1207–1216. doi:10.1002/hbm.20929

Allen, E.A., Erhardt, E.B., Damaraju, E., Gruner, W., Segall, J.M., Silva, R.F., Havlicek, M., Rachakonda, S., Fries, J., Kalyanam, R., Michael, A.M., Caprihan, A., Turner, J. a, Eichele,

ACCEPTED MANUSCRIPT

T., Adelsheim, S., Bryan, A.D., Bustillo, J., Clark, V.P., Feldstein Ewing, S.W., Filbey, F., Ford, C.C., Hutchison, K., Jung, R.E., Kiehl, K.A., Kodituwakku, P., Komesu, Y.M., Mayer, A.R., Pearlson, G.D., Phillips, J.P., Sadek, J.R., Stevens, M., Teuscher, U., Thoma, R.J., Calhoun, V.D., 2011. A baseline for the multivariate comparison of resting-state networks. Front. Syst. Neurosci. 5, 2. doi:10.3389/fnsys.2011.00002

Armento, M.E.A., Hopko, D.R., 2007. The Environmental Reward Observation Scale (EROS): Development, Validity, and Reliability. Behav. Ther. 38, 107–119. doi:10.1016/j.beth.2006.05.003

Barsaglini, A., Sartori, G., Benetti, S., Pettersson-Yeo, W., Mechelli, A., 2014. The effects of psychotherapy on brain function: A systematic and critical review. Prog. Neurobiol. 114, 1–4. doi:10.1016/j.pneurobio.2013.10.006

Beckmann, C.F., 2012. Modelling with independent components. Neuroimage 62, 891–901. doi:10.1016/j.neuroimage.2012.02.020

Bertha, E.A., Balázs, J., 2013. Subthreshold depression in adolescence: A systematic review. Eur. Child Adolesc. Psychiatry 22, 589–603. doi:10.1007/s00787-013-0411-0

Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default network: Anatomy, function, and relevance to disease. Ann. N. Y. Acad. Sci. 1124, 1–38. doi:10.1196/annals.1440.011

Calhoun, V. d., Adali, T., Pearlson, G. d., Pekar, J. j., 2001. A method for making group inferences from functional MRI data using independent component analysis. Hum. Brain Mapp. 14, 140–151. doi:10.1002/hbm.1048

Crone, E.A., Dahl, R.E., 2012. Understanding adolescence as a period of social–affective

Accepted manuscript

ACCEPTED MANUSCRIPT

engagement and goal flexibility. Nat. Rev. Neurosci. 13, 636–650. doi:10.1038/nrn3313

Cuijpers, P., Smit, F., Van Straten, A., 2007. Psychological treatments of subthreshold depression: A meta-analytic review. Acta Psychiatr. Scand. 115, 434–441. doi:10.1111/j.1600-0447.2007.00998.x

DeRubeis, R.J., Siegle, G.J., Hollon, S.D., 2008. Cognitive therapy vs. medication for depression: treatment outcomes and neural mechanisms. Nat. Rev. Neurosci. 9, 788–796. doi:10.1038/nrn2345

Euroqol Group, 1990. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy (New. York). 16, 199–208. doi:10.1016/0168-8510(90)90421-9

Faget-Agius, C., Boyer, L., Richieri, R., Auquier, P., Lançon, C., Guedj, E., 2016. Functional brain substrate of quality of life in patients with schizophrenia: A brain SPECT multidimensional analysis. Psychiatry Res. Neuroimaging. doi:10.1016/j.pscychresns.2016.02.005

Fair, D.A., Cohen, A.L., Dosenbach, N.U.F., Church, J.A., Miezin, F.M., Barch, D.M., Raichle, M.E., Petersen, S.E., Schlaggar, B.L., 2008. The maturing architecture of the brain's default network. Proc. Natl. Acad. Sci. 105, 4028–4032. doi:10.1073/pnas.0800376105

Franklin, G., Carson, A.J., Welch, K. a., 2015. Cognitive behavioural therapy for depression: systematic review of imaging studies. Acta Neuropsychiatr. 1–14. doi:10.1017/neu.2015.41

Fu, C.H.Y., Steiner, H., Costafreda, S.G., 2013. Predictive neural biomarkers of clinical response in depression: A meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. Neurobiol. Dis. doi:10.1016/j.nbd.2012.05.008

Goldapple, K., Segal, Z., Garson, C., Lau, M., Bieling, P., Kennedy, S., Mayberg, H., 2004.

Accepted manuscript

ACCEPTED MANUSCRIPT

Modulation of Cortical-Limbic Pathways in Major Depression: Treatment-Specific Effects of Cognitive Behavior Therapy. Arch. Gen. Psychiatry 61, 34–41.

Greicius, M.D., Flores, B.H., Menon, V., Glover, G.H., Solvason, H.B., Kenna, H., Reiss, A.L., Schatzberg, A.F., 2007. Resting-State Functional Connectivity in Major Depression: Abnormally Increased Contributions from Subgenual Cingulate Cortex and Thalamus. Biol. Psychiatry 62, 429–437. doi:10.1016/j.biopsych.2006.09.020

Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc. Natl. Acad. Sci. U. S. A. 100, 253–258. doi:10.1073/pnas.0135058100

Grimm, S., Ernst, J., Boesiger, P., Schuepbach, D., Hell, D., Boeker, H., Northoff, G., 2009. Increased self-focus in major depressive disorder is related to neural abnormalities in subcortical-cortical midline structures. Hum. Brain Mapp. 30, 2617–2627. doi:10.1002/hbm.20693

Hamilton, J.P., Etkin, A., Furman, D.J., Lemus, M.G., Johnson, R.F., Gotlib, I.H., 2012. Functional neuroimaging of major depressive disorder: A meta-analysis and new integration of baseline activation and neural response data. Am. J. Psychiatry 169, 693–703. doi:10.1176/appi.ajp.2012.11071105

Himberg, J., Hyvärinen, A., Esposito, F., 2004. Validating the independent components of neuroimaging time series via clustering and visualization. Neuroimage 22, 1214–1222. doi:10.1016/j.neuroimage.2004.03.027

Hwang, J.W., Xin, S.C., Ou, Y.M., Zhang, W.Y., Liang, Y.L., Chen, J., Yang, X.Q., Chen, X.Y., Guo, T.W., Yang, X.J., Ma, W.H., Li, J., Zhao, B.C., Tu, Y., Kong, J., 2016. Enhanced default

Accepted manuscript

ACCEPTED MANUSCRIPT

mode network connectivity with ventral striatum in subthreshold depression individuals. J. Psychiatr. Res. 76, 111–120. doi:10.1016/j.jpsychires.2016.02.005

Jacobson, N.S., Martell, C.R., Dimidjian, S., 2001. Behavioral Activation Treatment for Depression: Returning to Contextual Roots. Clin. Psychol. Sci. Pract. 8, 255–270. doi:10.1093/clipsy.8.3.255

Jinnin, R., Okamoto, Y., Takagaki, K., Nishiyama, Y., Yamamura, T., Okamoto, Y., Miyake, Y., Takebayashi, Y., Tanaka, K., Sugiura, Y., Shimoda, H., Kawakami, N., Furukawa, T.A., Yamawaki, S., 2017. Detailed course of depressive symptoms and risk for developing depression in late adolescents with subthreshold depression: A cohort study. Neuropsychiatr. Dis. Treat. 13, 25–33. doi:10.2147/NDT.S117846

Kanter, J.W., Mulick, P.S., Busch, A.M., Berlin, K.S., Martell, C.R., 2007. The Behavioral Activation for Depression Scale (BADS): Psychometric properties and factor structure. J. Psychopathol. Behav. Assess. 29, 191–202. doi:10.1007/s10862-006-9038-5

Kawakami, N., Takeshima, T., Ono, Y., Uda, H., Hata, Y., Nakane, Y., Nakane, H., Iwata, N., Furukawa, T.A., Kikkawa, T., 2005. Twelve-month prevalence, severity, and treatment of common mental disorders in communities in Japan: Preliminary finding from the World Mental Health Japan Survey 2002-2003. Psychiatry Clin. Neurosci. 59, 441–452. doi:10.1111/j.1440-1819.2005.01397.x

Kelly, C., Di Martino, A., Uddin, L.Q., Shehzad, Z., Gee, D.G., Reiss, P.T., Margulies, D.S., Castellanos, F.X., Milham, M.P., 2009. Development of anterior cingulate functional connectivity from late childhood to early adulthood. Cereb. Cortex 19, 640–57. doi:10.1093/cercor/bhn117

Accepted manuscript

ACCEPTED MANUSCRIPT

Kennedy, S.H., Konarski, J.Z., Segal, Z. V., Lau, M.A., Bieling, P.J., McIntyre, R.S., Mayberg, H.S., 2007. Differences in brain glucose metabolism between responders to CBT and venlafaxine in a 16-week randomized controlled trial. Am. J. Psychiatry 164, 778–788. doi:10.1176/appi.ajp.164.5.778

Kojima, M., Furukawa, A.T., 2003. Japanese Version of the Beck Depression Inventory, 2nd ed. Nippon. Co. Tokyo.

Konarski, J., Kennedy, S., Segal, Z., Lau, M., Bieling, P., McIntyre, R., Mayberg, H., 2009. Predictors of nonresponse to cognitive behavioural therapy or venlafaxine using glucose metabolism in major depressive disorder. J. psychiatry Neurosci. 34, 175–80.

Kunisato, Y., Takagaki, K., Okajima, I., Nakajima, S., Ishikawa, S., Kanai, Y., Okamoto, Y., Sakano, Y., Yamawaki, S., 2011. Development of Japanese version of Environmental Reward Observation Scale (EROS). Japanese J. Behav. Ther. 37, 21–31.

Lemogne, C., Mayberg, H., Bergouignan, L., Volle, E., Delaveau, P., Lehéricy, S., Allilaire, J.F., Fossati, P., 2010. Self-referential processing and the prefrontal cortex over the course of depression: A pilot study. J. Affect. Disord. 124, 196–201. doi:10.1016/j.jad.2009.11.003

Li, B., Liu, L., Friston, K.J., Shen, H., Wang, L., Zeng, L.L., Hu, D., 2013. A treatment-resistant default mode subnetwork in major depression. Biol. Psychiatry 74, 48–54. doi:10.1016/j.biopsych.2012.11.007

Lieberman, M.D., Cunningham, W.A., 2009. Type I and Type II error concerns in fMRI research: Re-balancing the scale. Soc. Cogn. Affect. Neurosci. 4, 423–428. doi:10.1093/scan/nsp052

Manoliu, A., Meng, C., Brandl, F., Doll, A., Tahmasian, M., Scherr, M., Schwerthöffer, D., Zimmer, C., Förstl, H., Bäuml, J., Riedl, V., Wohlschläger, A.M., Sorg, C., 2013. Insular

Accepted manuscript

ACCEPTED MANUSCRIPT

dysfunction within the salience network is associated with severity of symptoms and aberrant inter-network connectivity in major depressive disorder. Front. Hum. Neurosci. 7, 930. doi:10.3389/fnhum.2013.00930

Mason, L., Peters, E., Kumari, V., 2016. Functional connectivity predictors and mechanisms of cognitive behavioural therapies: A systematic review with recommendations. Aust. N. Z. J. Psychiatry 50, 311–21. doi:10.1177/0004867415624970

Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. Brain Struct. Funct. 1–13. doi:10.1007/s00429-010-0262-0

Nierenberg, A.A., Rapaport, M.H., Schettler, P.J., Howland, R.H., Smith, J.A., Edwards, D., Schneider, T., Mischoulon, D., 2010. Deficits in psychological well-being and quality-of-life in minor depression: Implications for DSM-V. CNS Neurosci. Ther. 16, 208–216. doi:10.1111/j.1755-5949.2009.00108.x

Nishimura, S., Tsuchiya, A., Hisashige, A., Ikegami, N., Ikeda, S., 1998. The development of the Japanese EuroQol Instrument. Heal. Care Soc. 8, 109–123.

Nolen-Hoeksema, S., 2000. The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. J. Abnorm. Psychol. 109, 504–11. doi:10.1037/0021-843X.109.3.504

Ordaz, S.J., Foran, W., Velanova, K., Luna, B., 2013. Longitudinal Growth Curves of Brain Function Underlying Inhibitory Control through Adolescence. J. Neurosci. 33, 18109–18124. doi:10.1523/JNEUROSCI.1741-13.2013

Pannekoek, J.N., Van Der Werff, S.J.A., Meens, P.H.F., Van Den Bulk, B.G., Jolles, D.D., Veer, I.M., Van Lang, N.D.J., Rombouts, S.A.R.B., Van Der Wee, N.J.A., Vermeiren, R.R.J.M.,

Accepted manuscript

ACCEPTED MANUSCRIPT

2014. Aberrant resting-state functional connectivity in limbic and salience networks in treatment-naive clinically depressed adolescents. J. Child Psychol. Psychiatry Allied Discip. 55. doi:10.1111/jcpp.12266

Pincus, H.A., Davis, W.W., McQueen, L.E., 1999. "Subthreshold" mental disorders. A review and synthesis of studies on minor depression and other "brand names". Br. J. Psychiatry 174, 288–296. doi:10.1192/bjp.174.4.288

Ritchey, M., Dolcos, F., Eddington, K.M., Strauman, T.J., Cabeza, R., 2011. Neural correlates of emotional processing in depression: Changes with cognitive behavioral therapy and predictors of treatment response. J. Psychiatr. Res. 45, 577–587. doi:10.1016/j.jpsychires.2010.09.007

Rubia, K., 2013. Functional brain imaging across development. Eur. Child Adolesc. Psychiatry 22, 719–731. doi:10.1007/s00787-012-0291-8

Rushworth, M.F.S., 2008. Intention, choice, and the medial frontal cortex. Ann. N. Y. Acad. Sci. doi:10.1196/annals.1440.014

Satomura, Y., Takizawa, R., Koike, S., Kawasaki, S., Kinoshita, A., Sakakibara, E., Nishimura, Y., Kasai, K., 2014. Potential biomarker of subjective quality of life: prefrontal activation measurement by near-infrared spectroscopy. Soc. Neurosci. 9, 63–73. doi:10.1080/17470919.2013.861359

Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., Gur, R.E., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. Neuroimage 60, 623–632. doi:10.1016/j.neuroimage.2011.12.063

Accepted manuscript

ACCEPTED MANUSCRIPT

Saverino, C., Grigg, O., Churchill, N.W., Grady, C.L., 2015. Age differences in the default network at rest and the relation to self-referential processing. Soc. Cogn. Affect. Neurosci. 10, 231–239. doi:10.1093/scan/nsu046

Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Reiss, A.L., Greicius, M.D., 2007. Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. J. Neurosci. 27, 2349–2356. doi:10.1523/JNEUROSCI.5587-06.2007

Shou, H., Yang, Z., Satterthwaite, T.D., Cook, P.A., Bruce, S.E., Shinohara, R.T., Rosenberg, B., Sheline, Y.I., 2017. Cognitive behavioral therapy increases amygdala connectivity with the cognitive control network in both MDD and PTSD. NeuroImage Clin. 14, 464–470. doi:10.1016/j.nicl.2017.01.030

Sundermann, B., Beverborg, M.O. lütke, Pfleiderer, B., 2014a. Meta-analysis of resting-state fMRI in depression: generating spatial hypotheses for potential clinical applications. PeerJ Prepr. 1–25. doi:10.7287/peerj.preprints.412v1

Sundermann, B., Olde Lütke Beverborg, M., Pfleiderer, B., 2014b. Toward literature-based feature selection for diagnostic classification: a meta-analysis of resting-state fMRI in depression. Front. Hum. Neurosci. 8, 692. doi:10.3389/fnhum.2014.00692

Takagaki, K., Okajima, I., Kunisato, Y., Nakajima, S., Kanai, Y., Ishikawa, S., Sakano, Y., 2013. Development and validation of the Japanese version of the Behavioral Activation for Depression Scale (BADS). Arch. Psychiatr. diagnostics Clin. Eval. 6, 76–85.

Takagaki, K., Okamoto, Y., Jinnin, R., Mori, A., Nishiyama, Y., Yamamura, T., Takebayashi, Y., Ogata, A., Okamoto, Y., Miyake, Y., Shimoda, H., Kawakami, N., Yamawaki, S., 2014. Behavioral characteristics of subthreshold depression. J. Affect. Disord. 168, 472–475.

Accepted manuscript

ACCEPTED MANUSCRIPT

doi:10.1016/j.jad.2014.07.018

Takagaki, K., Okamoto, Y., Jinnin, R., Mori, A., Nishiyama, Y., Yamamura, T., Yokoyama, S., Shiota, S., Okamoto, Y., Miyake, Y., Ogata, A., Kunisato, Y., Shimoda, H., Kawakami, N., Furukawa, A.T., Yamawaki, S., 2016. Behavioral activation for late adolescents with subthreshold depression: a randomized controlled trial. Eur. Child Adolesc. Psychiatry. doi:10.1007/s00787-016-0842-5

Yoshimura, S., Okamoto, Y., Onoda, K., Matsunaga, M., Okada, G., Kunisato, Y., Yoshino, A., Ueda, K., Suzuki, S., Yamawaki, S., 2014. Cognitive behavioral therapy for depression changes medial prefrontal and ventral anterior cingulate cortex activity associated with self-referential processing. Soc. Cogn. Affect. Neurosci. 9, 487–493. doi:10.1093/scan/nst009

Zeng, L.-L., Shen, H., Liu, L., Hu, D., 2014. Unsupervised classification of major depression using functional connectivity MRI. Hum. Brain Mapp. 35, 1630–1641. doi:10.1002/hbm.22278

Zeng, L.-L., Wang, D., Fox, M.D., Sabuncu, M., Hu, D., Ge, M., Buckner, R.L., Liu, H., 2014. Neurobiological basis of head motion in brain imaging. Proc. Natl. Acad. Sci. 111, 6058–6062. doi:10.1073/pnas.1317424111

Accepted manuscript

Table 1 Characteristics and the results of ANCOVAs for each clinical scores

ACCEPTED MANUSCRIPT

	Intervention group (n=19)		non-intervention group (n=21)			Group differences
	pre	post	pre	post	*t/F/χ²* values	*post-hoc comparison*
Age	18.21 (0.42)		18.19 (0.40)		0.15 *	
gender (f/ m)	6/13		7/14		0.01 n.s.	
BDI-II	15.42 (4.14)	9.26 (6.68)	15.38 (3.89)	14.38 (6.58)	7.34 *	intervention < non-intervention at post ** post < pre at intervention **
BADS-AC	10.58 (4.90)	18.63 (6.66)	15.33 (5.99)	14.86 (5.46)	26.00 **	intervention < non-intervention at pre ** pre < post at intervention **
EROS	22.11 (2.77)	25.42 (4.74)	22.38 (4.26)	22.19 (4.25)	6.67 *	non-intervention < intervention at post * pre < post at intervention **
EQ-5D total	0.90 (0.12)	0.90 (0.12)	0.86 (0.12)	0.84 (0.14)	0.14 n.s.	
VAS	65.05 (15.17)	75.68 (9.43)	66.19 (13.68)	58.90 (16.18)	12.23 **	non-intervention < intervention at post ** pre < post at intervention ** post < pre at non-intervention *

Note: BDI-II = Beck's Depression Inventory-II; BADS-AC = Behavioral Activation for Depression Scale-Activation; EROS = Environmental Reward Observation Scale; EQ-5D = EuroQOL-5 Dimension; VAS = visual analogue scale

n.s. $p > .10$, * $p < .05$, ** $p < .01$

Accepted manuscript

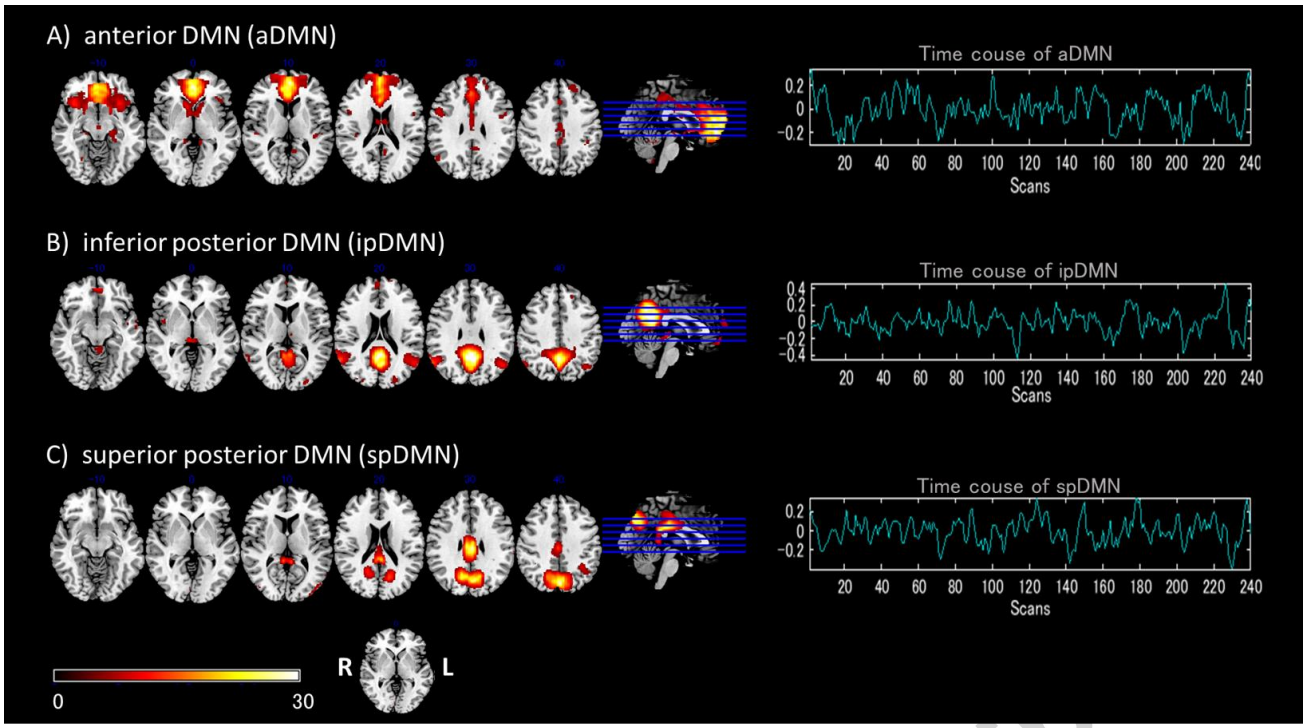
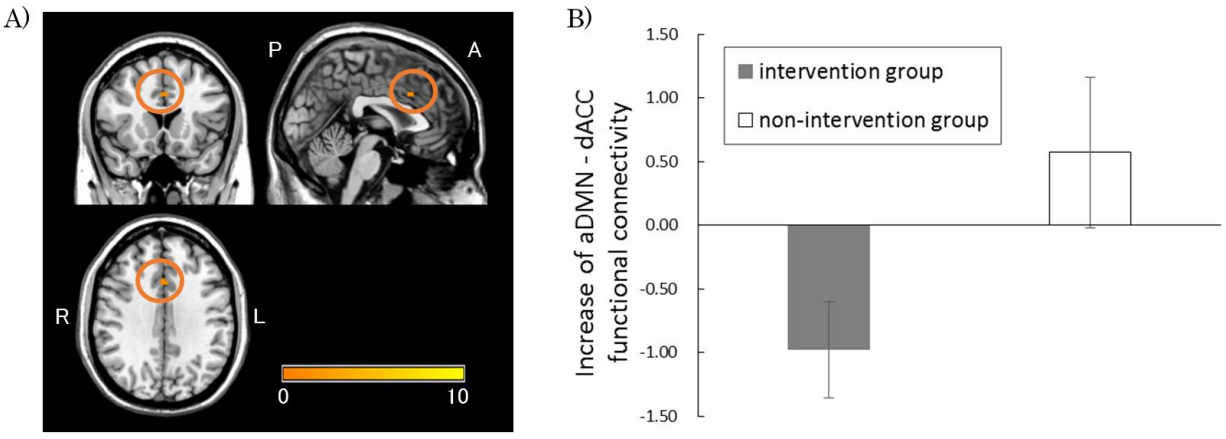
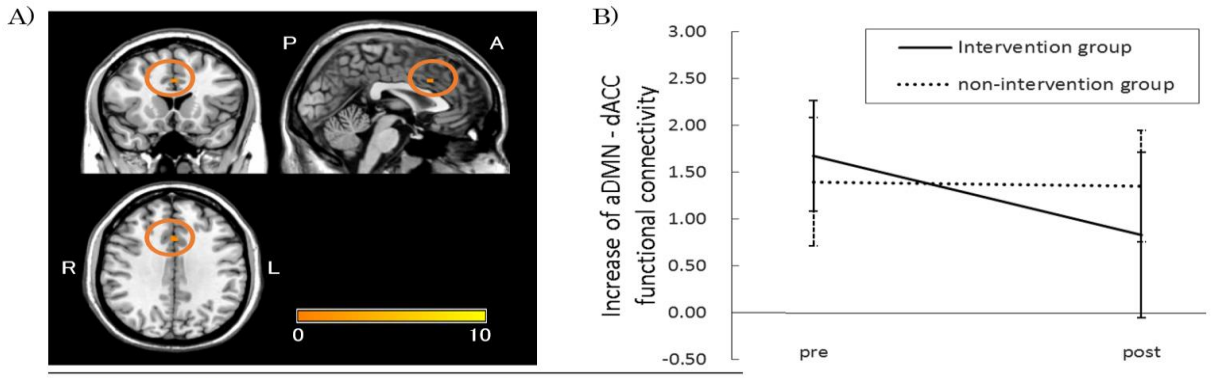


Fig. 1. T-maps and the corresponding time courses of three components reflecting the default mode subnetworks (DMN) in all participants (one-sample *t* test, $p < .05$ family-wise-error correction on the voxel-level and $p < .05$ false discovery rate correction on the cluster-level). (A) Anterior DMN (aDMN); (B) Inferior posterior DMN (ipDMN); (C) superior posterior DMN (spDMN). Color scales represent t-values from 0 to 30.

Accepted manuscript



BA	Side	*Z*	*p*	K_E	x	y	z
32	L/R	4.01	0.000	15	0	18	34



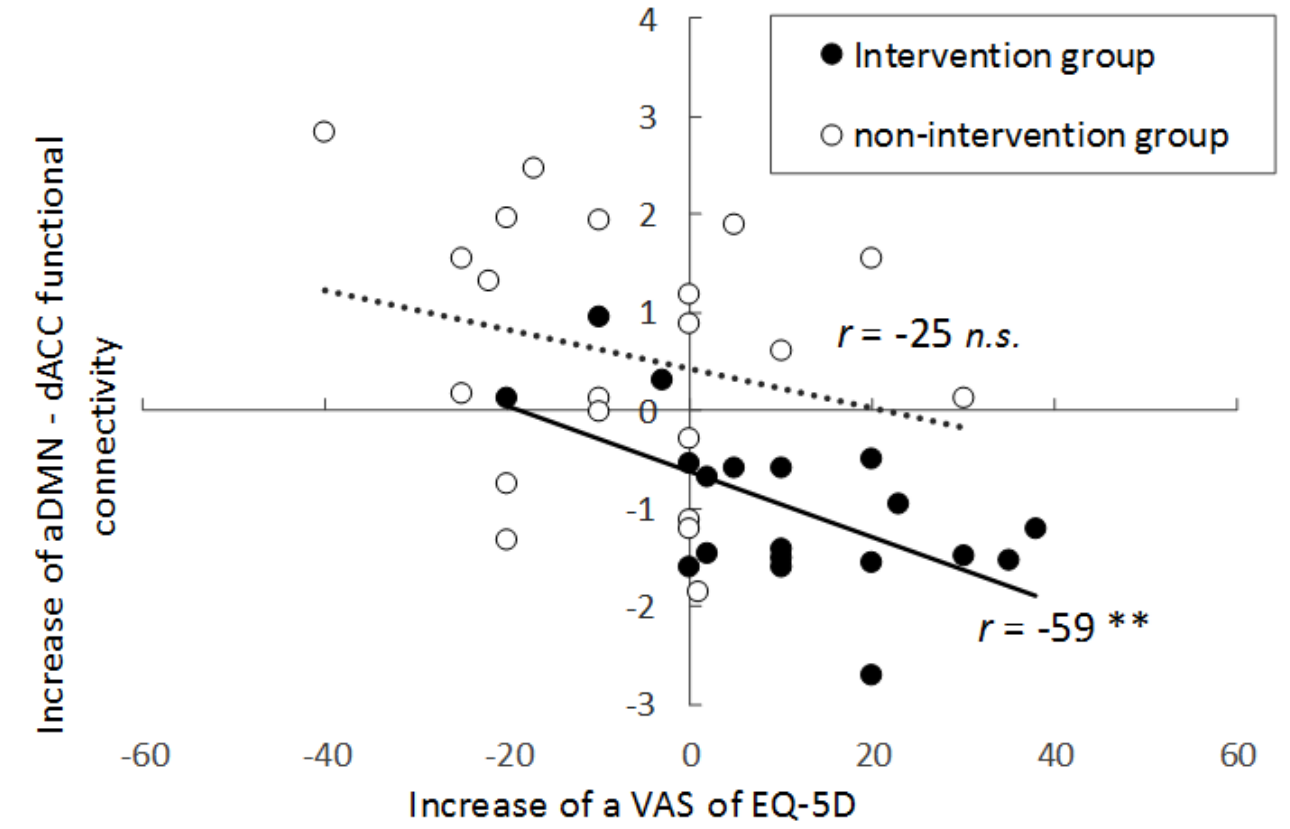
BA	Side	*Z*	*p*	K_E	x	y	z
32	L/R	3.98	0.000	10	0	18	34

Fig. 2. The dACC in the aDMN component that showed a significant interaction between group and time.

A) T-map of the activation greater or equal to 10 voxels reported at the threshold p =0.001, uncorrected. BA =Brodmann area; L =Left; R =Right; Z =Z value of the peak activation within the cluster; p = Uncorrected p value at peak level. Coordinates for the peak voxel are listed as MNI coordinates. A color scale represents t –value from 0 to 10. B) The change of dACC within the aDMN of each group. Error bars represent standard deviations.

aDMN = anterior default mode network; dACC = dorsal anterior cingulate.

Accepted



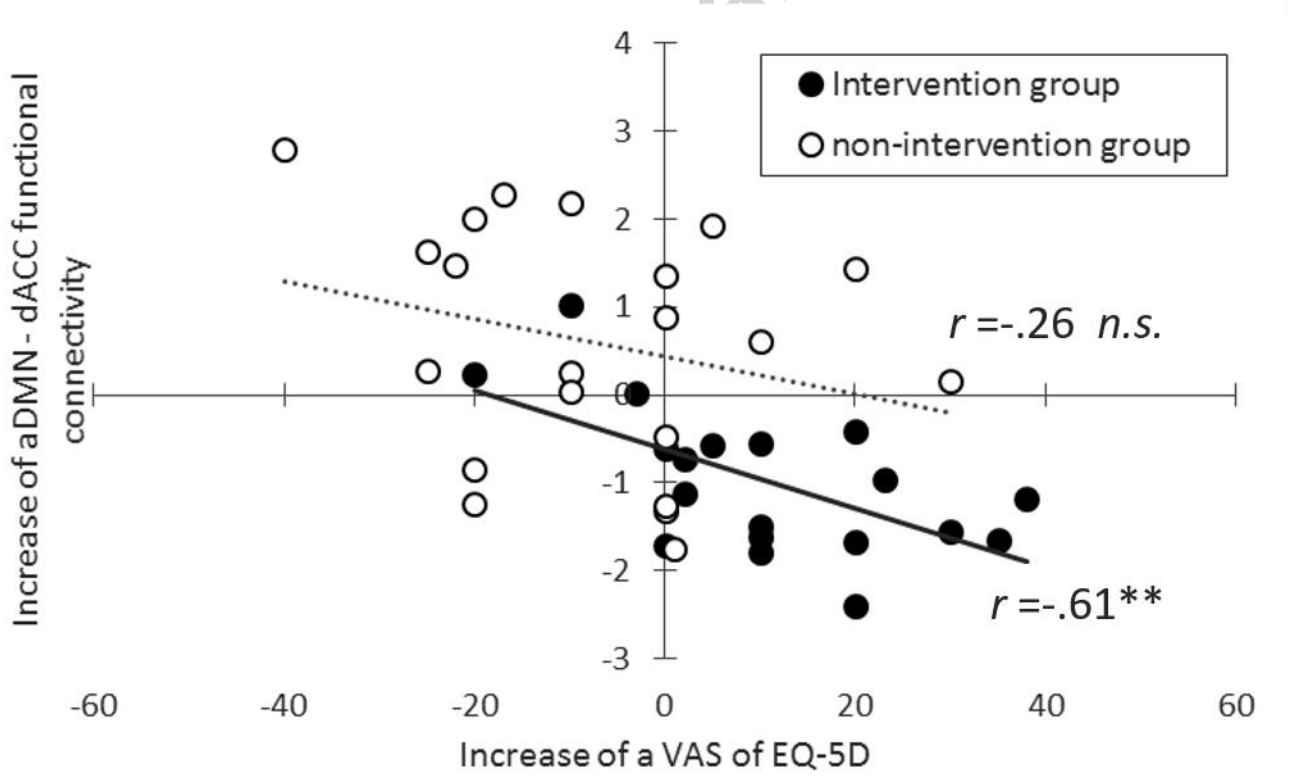


Fig. 3.

A scatter plot of correlations between the changes of aDMN-dACC functional connectivity and a

ACCEPTED MANUSCRIPT

VAS of EQ-5D in each group. Black = intervention group, and white= non-intervention group. aDMN =anterior default mode network; dACC = dorsal anterior cingulate cortex; VAS =visual analogue scale; EQ-5D =EuroQOL-5 Dimension.

$^{n.s.}$ $p > .10$, $^{**}p < .01$

Highlights

- Enhanced default mode network (DMN) has been demonstrated in subthreshold depressed person.
- Reduced DMN connectivity with the dorsal anterior cingulate was observed after behavioral activation.
- This reduction was correlated with an increase in health-related quality of life.

Accepted manuscript

Stress
The International Journal on the Biology of Stress

ISSN: 1025-3890 (Print) 1607-8888 (Online) Journal homepage: https://www.tandfonline.com/loi/ists20

Enriched environment ameliorates dexamethasone effects on emotional reactivity and metabolic parameters in mice

Eslen Delanogare, Raul Marin de Souza, Giovana Karoline Rosa, Fernando Garcia Guanabara, Alex Rafacho & Eduardo Luiz Gasnhar Moreira

To cite this article: Eslen Delanogare, Raul Marin de Souza, Giovana Karoline Rosa, Fernando Garcia Guanabara, Alex Rafacho & Eduardo Luiz Gasnhar Moreira (2020): Enriched environment ameliorates dexamethasone effects on emotional reactivity and metabolic parameters in mice, Stress

To link to this article: https://doi.org/10.1080/10253890.2020.1735344

Accepted author version posted online: 28 Feb 2020.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ists20

Check for updates

Full title: **Enriched environment ameliorates dexamethasone effects on emotional reactivity and metabolic parameters in mice.**

Running title: EE prevents dexamethasone behavior outcomes.

Eslen **Delanogare**[2], Raul Marin **de Souza**[2], Giovana Karoline **Rosa**[1], Fernando Garcia **Guanabara**[3], Alex **Rafacho**[1,4], Eduardo Luiz Gasnhar **Moreira**[1,2,4*.]

[1]Departamento de Ciências Fisiológicas, [2]Programa de Pós-Graduação em Neurociências, [3]Hospital Universitário Polydoro Ernani de São Thiago, Programa de Pós-Graduaão Multicêntrico em Ciências Fisiolóicas, Universidade Federal de Santa Catarina, 88040-900, Florianópolis, SC, Brasil.

*Corresponding author:

Eduardo Luiz Gasnhar Moreira, PhD (eduardo.luiz@ufsc.br); Departamento de Ciências Fisiológicas, Universidade Federal de Santa Catarina, 88040-900, Florianópolis, SC, Brasil. Phone: +55 48 3721 4617.

Accepted Manuscript

Abstract

Convincing evidence shows that stress is associated with the development and course of psychiatric and metabolic disorders. The hypothalamic-pituitary-adrenal (HPA) axis mediates the stress response, a cascade of events that culminate in the release of glucocorticoids from the adrenal cortex. Chronic hypercortisolism typically characterizes stress-related illnesses, such as depression, anxiety, and metabolic syndrome. Considering previous studies pointing that environmental enrichment (EE) mitigates the deleterious effects of stress on neurobiological systems, we hypothesized that EE can confer resiliency against prolonged glucocorticoid administration-induced behavioral and metabolic alterations in mice. In this regard, three-month-old male Swiss mice were exposed to a four-week period of standard environment (SE) or EE. After this period, still in the respective environments, dexamethasone was administered intraperitoneally (i.p.) at a dose of 4 mg/kg, for 21 consecutive days, in order to generate the emotional-related behavioral outcomes, as previously described. It is demonstrated herein that EE prevents the dexamethasone-induced anxiety-like and passive stress-coping behaviors, as observed in the open field and tail suspension tests. Moreover, EE mitigated the hyperproteinemia and body weight loss induced by excess dexamethasone and decreased basal glucose levels. Taken together, these results support the hypothesis that EE attenuates the effects of chronic administration of synthetic glucocorticoids in mice, a strategy that may be translated to the clinical perspective.

Keywords: dexamethasone; environmental enrichment; behavior dysfunctions, metabolic dysfunctions.

Accepted Manuscript

Introduction

The stress-responsive hypothalamic-pituitary-adrenal (HPA) axis is causally implicated in the development and course of major depressive disorder (MDD) (Keller et al., 2017), being described that patients with depression present HPA-axis overactivity (De Kloet, Joëls & Holsboer, 2005). The untoward consequences of excess cortisol are manifold, ranging from peripheral effects (e.g., glucose intolerance) to adverse effects on brain function. In this regard, hypercortisolemia, as seen in Cushing's disease or after prolonged glucocorticoid administration, can result in depression or emotional lability (Brown et al., 2006; Newell-Price et al., 2006). For instance, chronic dexamethasone administration in adult life (Skupio et al., 2015) or during intrauterine development (Conti et al., 2017) induces a range of depressive-like symptoms in rodents, such as anhedonia, behavioral despair, weight loss, and anxiety-like behavior.

The ability to cope with stress is influenced by multiple factors such as early life experiences, gender, or personality traits (Chen & Baram, 2016). In fact, many environmental factors seem to influence both the physiological functions of the central nervous system and its ability to counteract pathological changes. For instance, negative life events, such as the death of a parent, maternal deprivation, parental abandonment and separation of parents or divorce, can influence the development of mental disorders, such as depression (Juruena, 2013). In laboratory animals, stress reduction may be achieved by enriching the animal's environment with devices that promote its normal instinctive tendencies (Belz et al., 2003). In this regard, environmental enrichment (EE) is an animal housing technique composed of increased space, physical activity, and social interactions, which in turn increases sensory, cognitive, motor, and social stimulation (Nithianantharajah & Hannan, 2006). EE leads to improved cerebral health as defined by increased neurogenesis, enhanced learning and memory and resistance to external cerebral insults (Young et al., 1999). Moreover, living in an EE with physical and social stimulation also leads to improved metabolic health (Cao et al., 2011). Considering previous studies pointing that EE

Accepted Manuscript

mitigates the deleterious effects of stress on neurobiological systems and endocrine profiles (Mitra & Sapolsky, 2009; Lehmann & Herkenham, 2011), the present study hypothesized that EE would prevent the development of behavioral (e.g., depressive and anxiogenic-like effects) and metabolic alterations induced by chronic dexamethasone administration in mice. Thus, in order to evaluate the preventative effects of EE, mice were reared in this condition prior to the dexamethasone exposure.

Methods

Animals

Experiments were carried out on 2-month-old male Swiss albino mice from the animal facility of Universidade Federal de Santa Catarina (UFSC, Florianópolis, Brazil), weighing around 45-55 g at the time of testing. They were kept in groups of 10 animals per cage in a room under controlled temperature (23 ± 1 °C) and subjected to a 12-h light cycle (lights on 6:00 a.m.) with free access to food and water. All procedures comply with the guidelines on animal care of the local Ethics Committee on the Use of Animals (protocol 5497310718).

Drug administration

The synthetic glucocorticoid agonist dexamethasone (Laboratório Teuto, Brazil) was dissolved in saline and administered intraperitoneally (i.p.) at a dose of 4 mg/kg of body weight for 21 consecutive days. The dose of dexamethasone was chosen based on the study by Skupio and colleagues (2015), which demonstrated that chronic dexamethasone administration induces both a range of depressive-like symptoms, such as behavior despair, weight loss, and anxiety-like behavior. The control group received saline injection (0.1 mL per 10 grams of body weight). The treatments occurred between 7:00 and 9:00 a.m.

Accepted Manuscript

Housing Conditions

The standard environment (SE) mice were housed in wire-topped clear plastic cages (38 cm × 32 cm × 17 cm). The EE mice were housed in larger plastic boxes (44 cm × 32 cm × 18 cm), containing a variety of stimuli, e.g., a running wheel, plastic tubing, ladders, rubber balls and a wood shelter, as previously established in or laboratory (de Souza et al., 2019a). These items were moved to different spatial locations of the cage every 3-4 days. The animals were maintained 24 h per day in groups of 10 animals for both SE and EE conditions until the beginning and throughout the duration of the behavioral experiments. Although in several studies the mice designated to the EE cohort were housed in larger groups to allow more social interactions, we kept the same number of animals per cage, as in previous studies of our research group (de Souza et al., 2019a).

Experimental design

Forty male Swiss adult mice were divided into four experimental groups (10 mice per group): standard environment plus saline (SE Saline), standard environment plus dexamethasone (SE Dexa), environmental enrichment plus saline (EE Saline) and environmental enrichment plus dexamethasone (EE Dexa). Animals were exposed to a 28-days period of SE or EE. After this period, still in the respective environments, the animals received daily saline solution or dexamethasone, as previously described, for 21 consecutive days. Body mass was measured daily. On days 47 and 48, still on the influence of the above-mentioned treatments, the animals were submitted to two behavioral tests: tail suspension test (day 47) and open field test (day 48). One day after the last behavioral test, mice were food-deprived for six hours (starting 7 a.m.) and the glucose tolerance test was performed (day 49). A day after (day 50), mice were again food-deprived for six hours, anesthetized with a mixture of ketamine (80 mg/kg, i.p.) and xylazine (10 mg/kg, i.p.), and the blood was collected from the heart to determination of plasma

Accepted Manuscript

triglycerides, cholesterol, total protein, and corticosterone levels. On day 50, no dexamethasone was administered (Figure 1).

Behavioral tasks

All experiments were conducted during the light phase (1:00–6:00 p.m.). Experiments were conducted in a controlled-temperature room (23°C, the humidity was between 40% - 60%), with low-light intensity (12 lx). Moreover, they were recorded (webcam Microsoft VX 3000) and the videos were analyzed using the Ethowatcher® software.

Tail Suspension Test (TST)

The TST measures coping strategy to an acute inescapable stress (Commons et al., 2017). This test was developed by Steru and colleagues (1985) based on the premise that an animal subjected to a stressful and inescapable situation presents two types of alternating behaviors, the agitation characteristic of the attempt to escape from the stressful situation, and immobility. This pattern of behavior can also be called searching-behavior, characterized by the alternation of intense motor activity and energy expenditure with immobility. Mice both acoustically and visually isolated were suspended 30 cm above the floor by adhesive tape placed approximately 1 cm from the tip of the tail. Immobility time was manually recorded during a 6-min period by an experienced observer. The observer was in the room where experiments were performed and were blind to the animal condition.

Accepted Manuscript

Open Field Test (OFT)

The open field was used to evaluate the locomotor and exploratory activities induced by a novel environment. The natural tendency of the animal in a new environment is to exploit it, despite the stress and conflict caused by the new environment (Prut & Belzung, 2003). This test

assess unconditioned emotional reactions based on natural approach/avoidance conflict (Ramos, 2008). The anxiety-related behavior in the OFT is triggered by two factors: individual testing and agoraphobia. Higher levels of anxiety should mainly lead to decreases in the ratio 'number of squares visited in center/number of squares visited on periphery'. The apparatus, made of wood and covered with impermeable Formica, had 50 cm wide × 50 cm deep × 40 cm high. Each mouse was placed in the center of the open field and allowed to freely explore the apparatus for 5 min. The following behavioral parameters were evaluated: % of crosses in the center, total number of crosses and time in the center.

Biochemical analysis

The glucose tolerance test was assessed by i.p. administration of D-(+)-glucose (Sigma Aldrich, St Louis, MO) at a concentration of 2 g/kg body weight. Blood glucose was measured from the tail tip at baseline and 15, 30, 60, and 120 minutes after glucose overload using a glucometer (Accu-Chek Performa®). For the area under the curve (AUC) calculation, it was used the mean baseline value (time 0) for each group (baseline for the area). Total cholesterol, triglycerides, and total proteins were measured in plasma using the enzymatic kit according to the manufacturer's instructions (Gold Analisa, Belo Horizonte, MG, Brazil). The results were expressed as mg/dl.

Statistical analysis

The results are presented as mean + standard error of the mean (SEM). Statistical analyses were carried out using a two-way or three-way analysis of variance (ANOVA). Following significant ANOVAs, multiple post hoc comparisons were performed using Newman Keul's test. P values less than 0.05 ($P < 0.05$) were considered as indicative of significance. All tests were performed using the STATISTICA® software package (StatSoft Inc, Tulsa, OK, USA).

Accepted Manuscript

Results

Open Field Test (OFT)

The two-way ANOVA indicated a significant main effect for treatment [$F(1, 36) = 5.18$, $p < 0.05$] and for housing conditions by treatment interaction [$F(1, 36) = 6.93$, $p < 0.05$] on center time (Fig 2a). Subsequent post hoc comparisons revealed a significant decreased in the time spent in the center of the apparatus in the SE Dexa group in comparison with the SE Saline group and with the EE Dexa group ($p < 0.05$). On the other hand, the two-way ANOVA indicated no significant effects on the percentage of center crossings (Fig 2b) and on the total crossings (Fig 2c).

Tail Suspension Test (TST)

The two-way ANOVA indicated no significant effects on the latency to immobility (Fig 3a). On the other hand, the two-way ANOVA indicated a significant main effect for treatment [$F(1, 36) = 8.58$, $p < 0.05$] and for housing conditions by treatment interaction [$F(1, 36) = 9.65$, $p < 0.005$] on the immobility time in the TST. Subsequent post-hoc analysis revealed a significant increase in the immobility time in the SE Dexa group when compared with the SE Saline group and with the EE Dexa group ($p < 0.05$) (Fig 3b).

Metabolic analysis

The two-way ANOVA revealed a significant main effect for treatment [$F(1, 36) = 27.98$, $p < 0.0005$] and for housing conditions by treatment interaction [$F(1, 36) = 4.06$, $p < 0.05$] on body mass variation (Fig 4a). Subsequent post-hoc analysis revealed a significant loss of body mass (%) in the SE Dexa group when compared with the SE saline group and with the EE Dexa group ($p < 0.05$).

The two-way ANOVA analysis indicated a significant main effect for treatment [$F(1, 36) = 153.86$, $p < 0.0005$] on plasma cholesterol concentration (Fig 4b). Subsequent post-hoc analysis

Accepted Manuscript

revealed a significant elevation on plasma cholesterol concentrations in both the SE Dexa group and the EE Dexa group when compared with their respective saline control groups ($p < 0.05$). On the other hand, the two-way ANOVA indicated a significant main effect for treatment [$F(1, 36) = 5.12$, $p < 0.05$] and for housing conditions by treatment interaction [$F(1, 36) = 6.26$, $p < 0.05$] on plasma triglyceride concentration (Fig 4c). Subsequent post-hoc analysis revealed a significant decrease in triglycerides concentrations in the SE Dexa group when compared with the SE Saline group and with the EE Dexa group ($p < 0.05$). Moreover, the two-way ANOVA revealed a significant main effect for treatment [$F(1, 36) = 5,85$, $p < 0.05$] and for housing conditions by treatment interaction [$F(1, 36) = 5,61$, $p < 0.05$] on plasma total protein concentration. Subsequent post-hoc analysis revealed a significant increase in total protein concentration in the SE Dexa group when compared with the SE Saline group and with the EE Dexa group ($p < 0.05$) (Fig 4d).

The two-way ANOVA analysis indicated a significant main effect for housing conditions [$F(1, 36) = 10.72$, $p < 0.005$] on basal glucose levels (Fig 4e). Subsequent post-hoc analysis revealed a significant decrease on basal glucose levels specifically in the EE Dexa group when compared with its respective saline control group ($p < 0.05$). The three-way ANOVA indicated significant main effects for housing conditions [$F(1, 180)= 11.68$, $p < 0.001$], treatment [$F(1, 180)= 4.81$, $p < 0.05$], time [$F(1, 180)= 11.68$, $p < 0.0001$], and housing conditions by treatment interaction [$F(1, 180)= 11.35$, $p < 0.0005$] (Fig 4f). Subsequent post-hoc comparisons indicated no significant differences among groups at each time point. On the other hand, the two-way ANOVA revealed a significant main effect for treatment [$F(1, 36) = 5.81$, $p < 0.05$] on the area-under-the-glucose-curve (AUC) (Fig 3g) that was based on blood glucose values from Figure 4g.

Accepted Manuscript

Discussion

The present study aimed to investigate the effects of EE on dexamethasone-induced behavioral reactivity and metabolic disturbances in mice. Our findings that chronic dexamethasone treatment induces anxiety-like and passive stress-coping behaviors in mice replicate the findings of the study conducted by Skupio et al. (2015), upon which our experimental protocol was based. Regarding the metabolic effects of dexamethasone, mice presented weight loss, decreased plasma triglycerides levels, increased total plasma proteins, and cholesterol levels and glucose intolerance. Of importance, we demonstrated that EE blunted the dexamethasone-induced behavioral reactivity in the TST and OFT, respectively. EE also mitigated dexamethasone-induced weight loss, hypotriglyceridemia, and hyperproteinemia.

Herein, chronic dexamethasone treatment-induced behavioral despair and anxiety-like behavior in mice, as observed in the TST and OFT, respectively. These findings agree with data observed in the literature. There is compelling evidence that dexamethasone administration, in an acute or chronic regimen, at doses varying from 16 µg/kg to 4 mg/kg, induces depressive-like and anxiety-like behavior in mice (Wróbel et al., 2014; Skupio et al., 2015; De Souza et al., 2019b). Moreover, Pazini et al. (2017) have shown that swiss mice treated with corticosterone, at a dose of 20 mg/kg, for 21 days, showed an increase in the immobility time in the TST. The authors associated such depressive-like effect with decreased hippocampal cell proliferation and neuronal differentiation and increased glial fibrillary acid protein (GFAP) immunostaining (suggestive of astrogliosis) in dentate gyrus (DG) of the hippocampus (Pazini et al., 2017). Moreover, Zhang and colleagues have shown that C57Bl/6 mice treated with corticosterone, at a dose of 40 mg/kg, for 35 days showed significantly less time spent in the center zone of the OFT (Zhang et al., 2016). The behavioral despair and anxiety-like behavior seen here probably reflect the spectrum of mood disorders observed in patients with hypercortisolemia, as in Cushing's syndrome (Newell-Price et al., 2006) and the common comorbidity of depression and anxiety observed in patients (Wu &

Accepted Manuscript

Fang, 2014). In the present study, it was found that treatment with EE mitigated both behavioral despair and anxiety-like behavior induced by dexamethasone in the TST and OFT, respectively. In this regard, previous studies reported that EE decreases corticosterone concentrations in rodents (Belz et al., 2003). In an elegant study, Sztainberg and colleagues (2010) demonstrated that EE significantly reduced CRHr-1 mRNA expression in limbic regions, suggesting a molecular mechanism that might explain the EE effects on decreasing corticosterone concentrations: the downregulation of this receptor decreases the release of ACTH by adenohypophysis and, consequently, decreases the release of corticosterone by the adrenal cortex. Furthermore, Hare and colleagues (2014) showed that four weeks of voluntary wheel running resulted in a shorter time to peak corticosterone concentration and a more rapid decay of plasma corticosterone levels following restrain stress in mice. The authors conclude that exercise renders animals particularly resilient, capable of producing a robust response to stress while limiting the damage that might result from overexposure to glucocorticoids during periods of stress (Hare et al., 2014). Indeed, numerous clinical and experimental studies demonstrated that many environmental factors, such as EE, may affect both the physiological functions of the central nervous system and its ability to counteract pathological changes (Van Praag, Kempermann & Gage, 2000). Of note, the changes provided following EE exposure are commonly described as experience-dependent, indicating that they result from active interaction between the animal and the affordances available in the environment (Nithianantharajah & Hannan, 2006).

Herein it was observed that chronic dexamethasone treatment induced a significant metabolic dysfunction, characterized by glucose intolerance, decreased body weight and plasma triglyceride levels, as well as increased plasma total protein and cholesterol levels. The increase in plasma total protein concentrations is related to the catabolic properties of glucocorticoids, i.e., increasing protein degradation and decreasing protein synthesis in skeletal muscle (Morgan et al., 2016), which may also support the weight loss observed in dexamethasone-treated rodents (Tonolo et al.,

Accepted Manuscript

1988; Noh et al., 2014). This reduction in body weight during dexamethasone treatment is also related to a negative water balance (Thunhorst et al., 2007) and higher levels of anorexigenic insulin levels (Protzek et al., 2014). The dexamethasone treatment also induced a significant decrease in plasma triglyceride concentration. However, in this case, our results diverge from prior findings (Barbosa et al., 2016). For instance, dexamethasone treatment, at a dose of 0.5 mg/kg for 15 days, resulted in a significant increase in triglyceride levels in rats (Barbosa et al., 2016). Although most data have shown that glucocorticoids induce lipolysis and consequently, plasma free fatty acid increases (Divertie et al., 1991), the effects of glucocorticoids on lipid mobilization are still controversial. For instance, some studies report an inhibitory effect of cortisol on lipolytic activity (Ottosson et al., 2000). Of note Chen and colleagues (2018) found a significant decrease of triglyceride in plasma and in the liver of goats after intramuscular administration of dexamethasone, at a concentration of 0.2 mg/kg for 21 days. Finally, the significant increase in plasma cholesterol concentration here observed is in accordance with previous studies. For instance, Kumar and colleagues (2011) showed that dexamethasone administration, at a dose of 10 mg/kg for 8 days, increased the plasma cholesterol levels in rats. Most importantly herein, we demonstrated that EE mitigated the dexamethasone effects on body weight, plasma triglycerides, and total protein levels. EE has a general tendency to increase activity and activity energy expenditure in rodents, which has downstream effects on other aspects of energy balance (Novak et al., 2012). In this regard, it is noteworthy that our EE protocol included a running wheel, and since physical exercise is associated with increased protein synthesis (Tipton & Wolfe, 2001), one may associate with the maintenance of body mass observed in EE dexamethasone-treated mice. After ~ 3 weeks in running well, skeletal muscles (gastrocnemius, tibialis anterior, triceps) of male mice show enhanced citrate synthase activity, a validated biomarker for mitochondrial density in skeletal muscle, resulting in better endurance capacity (Manzanares et al., 2018). Moreover, Barel

Accepted Manuscript

and colleagues showed that run on a treadmill attenuated the dexamethasone-induces muscular glycogen loss, muscular atrophy and body weight loss in rats (Barel et al., 2010).

Glucocorticoids, *per se*, are considered diabetogenic agents, as they may promote an increase in hepatic glucose production and a decrease in peripheral uptake (e.g., in muscle and adipose tissues) (Pasieka and Rafacho, 2016). Because of such properties, up to 60% of patients with Cushing's syndrome, characterized by hypercortisolemia, have glucose intolerance (Newell-Price et al., 2006). The glucose intolerance caused by dexamethasone treatment in our study is in agreement with previous studies (Pandey et al., 2014; Lee et al., 2018). The reduction of insulin-stimulated protein kinase B phosphorylation in skeletal muscle seems to be consensual at this chronic dexamethasone exposure (Pandey et al., 2014, Lee et al., 2018), which can imply in reduced glucose disposal. Despite to be not effective in preventing the glucose intolerance caused by glucocorticoids, EE significantly improved the basal glycemic values and kept blood glucose levels bellow of that observed in dexamethasone-treated mice in SE during all period of the test. In this regard, McMurphy and colleagues (2018) observed that EE exposure induces a robust reduction of brown and white adipose tissue, in addition to decreasing hepatic steatosis, hepatic glucose production and increasing glucose uptake by the liver and adipose tissue, contributing to glycemic control in mice. Specifically, the authors demonstrated that EE induced a significant reduction in the expression of two important enzymes involved in the gluconeogenesis pathway, phosphoenolpyruvate carboxykinase 1 (PCK1) and glucose-6-phosphatase (G6PC) (McMurtphy et al., 2018). The apparent lack of effect of EE upon beta cell function may be due to the genomic impact of dexamethasone on insulin secretory machinery of beta cells (Lambillotte et al., 1997). Although we do not discard new approaches of EE aiming to be more effective, we consider EE as a valuable tool for improvement of neurometabolic outcomes as previously demonstrated in a model of glucose intolerance associated to hypercholesterolemic diet where EE mitigated the glucose intolerance (De Souza et al., 2019a).

Accepted Manuscript

An inevitable limitation of our study is that multiple sensory, social, and physical activity factors are typically combined in the EE treatment, making it difficult to determine which of the various enrichment factors or the interaction contributes to the observed effects on behavioral and metabolic outcomes. In this regard, it is not possible to distinguish between the effects of EE *vs* exercise alone, since a running wheel was provided. Of note, McMurphy and colleagues (2018) evaluated the extent to which physical activity is the responsible for the EE-induced healthy aging. In that study, the gene expression profile of the major organs involved in systemic glucose homeostasis (liver, fat and muscle) revealed distinct patterns between EE and running wheel animals, where animals exposed to running wheel showed minor changes in these parameters (McMurphy et al., 2018). Thus, it is noteworthy that no social interaction, novelty, or any other single variable can account for all of the effects of EE (van Praag et al., 2000).

Conclusion

Collectively, our results reproduce the depressive- and anxiety-like behavior caused by supraphysiologic doses of dexamethasone in Swiss male mice, as well as metabolic disturbances (glucose intolerance, hyperproteinemia, hypercholesterolemia, and body weight loss). Most importantly, EE mitigates the depressive- and anxiety-like behaviors. Moreover, our findings suggest that EE is able to promote healthier metabolic function in mice, even in situations of chronic administration of synthetic glucocorticoids. It is noteworthy, however, that although our results reinforce the positive infuences of EE on stress response, the interpretation of our data should be seen as prophylactic rather than therapeutic since EE started before the dexamethasone treatment. Taken together, our results support the hypothesis that EE enhances resilience to stressors, such as the chronic administration of synthetic glucocorticoids used herein, a strategy that may be translated to the clinical perspective. Both, behavioral and metabolic improvements, are of particularly importance, since the prevalence of metabolic disturbance is 58% higher in

Accepted Manuscript

psychiatric patients than in the general population, which suggests that metabolic disturbance is a general comorbidity seen in different psychiatric patient groups, including patients with major depressive disorder (Vancampfort et al., 2015) and anxiety disorder (Tang et al., 2017).

Acknowledgments

Research supported by grants from the Brazilian funding agencies: CAPES, CNPq (Universal 424799/2018-9) and FAPESC.

Declaration of interest

The authors have no financial or personal conflicts of interest related to this work.

References

Abdelmannan, D., Tahboub, R., Genuth, S., & Ismail-Beigi, F. (2010). Effect of dexamethasone on oral glucose tolerance in healthy adults. *Endocr Pract.* 16(5):770-7.

Barbosa, A. M., Francisco, Pde. C., Motta, K., Chagas, T. R, Dos Santos, C., Rafacho, A., & Nunes, E. A. (2016). Fish oil supplementation attenuates changes in plasma lipids caused by dexamethasone treatment in rats. *Appl Physiol Nutr Metab*. 41(4):382-90.

Barel, M., Perez, O. A., Giozzet, V. A., Rafacho, A., Bosqueiro, J. R, & do Amaral, S. L. (2010). Exercise training prevents hyperinsulinemia, muscular glycogen loss and muscle atrophy induced by dexamethasone treatment. *Eur J Appl Physiol*, 108(5):999-1007.

Belz, E. E., Kennell, J. S., Czambel, R. K., Rubin, R. T & Rhodes, M. E. (2003). Environmental enrichment lowers stress-responsive hormones in singly housed male and female rats. *Pharmacol Biochem Behav*., 76(3-4):481-6.

Brown, E. S., Vera, E., Frol, A. B., Woolston, D. J., & Johnson, B. (2006). Effects of chronic prednisone therapy on mood and memory. *Journal of affective disorders*, 99(1-3), 279–283.

Cao, L., Choi, E. Y., Liu, X., Martin, A., Wang, C., Xu, X., & During, M. J. (2011). White to brown fat phenotypic switch induced by genetic and environmental activation of a hypothalamic-adipocyte axis. *Cell metabolism*, 14(3), 324–338.

Chen, Q., Niu, L., Hua, C., Geng, Y., Cai, L., Tao, S., Ni, Y & Zhao R. (2018) Chronic dexamethasone exposure markedly decreased the hepatic triglyceride accumulation in growing goats. *Gen Comp Endocrinol*. 1;259:115-121.

Accepted Manuscript

Chen, Y., & Baram, T. Z. (2016). Toward Understanding How Early-Life Stress Reprograms Cognitive and Emotional Brain Networks. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, 41(1), 197–206.

Commons, K.G., Cholanians, A. B., Babb, J. A., Ehlinger, D. G. (2017) The rodent forced swin test measures stress-coping strategy, not depression-like behavior. *ACS Chem Neurosci*. 8(5): 955-960.

Conti, M., Spulber, S., Raciti, M & Ceccatelli S. (2017) Depressive-like phenotype induced by prenatal dexamethasone in mice is reversed by desipramine. *Neuropharmacology*, 126:242-249.

de Kloet, E. R., Joëls, M & Holsboer, F. (2005). Stress and the brain: from adaptation to disease. *Nat Rev Neurosci*, 6(6):463-75.

de Souza, R.M., de Souza, L., Machado, A.E., de Bem Alves. A.C., Rodrigues, F.S., Aguiar, A.S. Jr., Dos Santos, A.R.S., de Bem, A.F., & Moreira, E.L.G. (2019a). Behavioural, metabolic and neurochemical effects of environmental enrichment in high-fat cholesterol-enriched diet-fed mice. *Behav Brain Res*, 1;359:648-656.

de Souza, I. B. M. B., Costa, L. R. F., Tiago, P. R. F., Cagni, F. C., Lima, R. H, Silva Junior, E. D., & Gavioli, E. C. (20019b). Venlafaxine and nortriptyline reverse acute dexamethasone-induced depressive-like behaviors in male and female mice. *Exp Clin Psychopharmaco*, 27(5):433-442.

Delaunay, F., Khan, A., Cintra, A., Davani, B., Ling, Z. C., Andersson, A., … Okret, S. (1997). Pancreatic beta cells are important targets for the diabetogenic effects of glucocorticoids. *The Journal of clinical investigation*, *100*(8): 2094–2098.

Divertie, G. D., Jensen, M. D & Miles, J. M. (1991). Stimulation of lipolysis in humans by physiological hypercortisolemia. *Diabetes*, 40 1228–1232.

Djurhuus, C. B., Gravholt, C. H., Nielsen, S., Mengel, A., Christiansen, J. S., Schmitz, O. E., & Møller, N. (2002). Effects of cortisol on lipolysis and regional interstitial glycerol levels in humans. *Am J Physiol Endocrinol Metab*. 283(1):E172-7.

Hare, B. D., Beierle, J. A., Toufexis, D. J., Hammack, S. E., & Falls, W. A. (2014). Exercise-associated changes in the corticosterone response to acute restraint stress: evidence for increased adrenal sensitivity and reduced corticosterone response duration. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, 39(5), 1262–1269.

Juruena, M. F. (2013). Early-life stress and HPA axis trigger recurrent adulthood depression. *Epilepsy & Behavior*, 38, 148–159.

Keller, J., Gomez, R., Williams, G., Lembke, A., Lazzeroni, L., Murphy, G. M., Jr, & Schatzberg, A. F. (2017). HPA axis in major depression: cortisol, clinical symptomatology and genetic variation predict cognition. *Molecular psychiatry*, 22(4), 527–536.

Kumar, V. R., Inamdar, M. N., Nayeemunnisa, & Viswanatha, G. L. (2011). Protective effect of lemongrass oil against dexamethasone induced hyperlipidemia in rats: possible role of decreased lecithin cholesterol acetyl transferase activity. *Asian Pac J Trop Med*. 4(8):658-60.

Accepted Manuscript

Lambillotte, C., Gilon, P., & Henquin, J. C. (1997). Direct glucocorticoid inhibition of insulin secretion. An in vitro study of dexamethasone effects in mouse islets. *The Journal of clinical investigation*, 99(3), 414–423.

Lee, M. K., Choi, J. W., Choi, Y. H., & Nam, T. J. (2018). *Pyropia yezoensis* Protein Supplementation Prevents Dexamethasone-Induced Muscle Atrophy in C57BL/6 Mice. *Marine drugs*, *16*(9), 328.

Lehmann, M. L,. & Herkenham, M. (2011). Environmental enrichment confers stress resiliency to social defeat through an infralimbic cortex-dependent neuroanatomical pathway. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *31*(16), 6159–6173.

McMurphy, T., Huang, W., Queen, N. J., Ali, S., Widstrom, K. J., Liu, X., Cao, L. (2018). Implementation of environmental enrichment after middle age promotes healthy aging. *Aging*, 10(7): 1698–1721.

Manzanares, G., Brito-da-Silva, G., & Gandra, P. G. (2018). Voluntary wheel running: patterns and physiological effects in mice. *Brazilian journal of medical and biological research*, 52(1), e7830.

Mitra, R., Sapolsky R.M. (2009). Effects of enrichment predominate over those of chronic stress on fear-related behavior in male rats. *Stress,* 12:4, 305-312,

Morgan, S. A., Hassan-Smith, Z. K., Doig, C. L., Sherlock, M., Stewart, P. M., & Lavery, G. G. (2016). Glucocorticoids and 11β-HSD1 are major regulators of intramyocellular protein metabolism. *The Journal of endocrinology*, 229(3), 277–286.

Newell-Price, J., Bertagna, X., Grossman, Ab & Nieman, Lk. (2006) Cushing's syndrome. *Lancet,* 13;367(9522):1605-17.

Nicolaides, N. C., Charmandari, E., Kino, T., & Chrousos, G. P. (2017). Stress-Related and Circadian Secretion and Target Tissue Actions of Glucocorticoids: Impact on Health. *Frontiers in endocrinology*, 8, 70.

Nithianantharajah, J & Hannan, A. J. (2006). Enriched environments, experience dependent plasticity and disorders of the nervous system. *Nat Rev Neurosci*. 2006;7:697.

Noh, K. K., Chung, K. W., Choi, Y. J., Park, M. H., Jang, E. J., Park, C. H., … Chung, H. Y. (2014). β-Hydroxy β-methylbutyrate improves dexamethasone-induced muscle atrophy by modulating the muscle degradation pathway in SD rat. *PloS one*, 9(7), e102947.

Novak, C. M., Burghardt, P. R., Levine, J. A. (2012). The use of a running wheel to measure activity in rodents: relationship to energy balance, general activity, and reward. *Neurosci Biobehav Rev*, 36:1001-14.

Ottosson, M., Lönnroth, P., Björntorp, P & Edén, S. (2000). Effects of cortisol and growth hormone on lipolysis in human adipose tissue. *J Clin Endocrinol Metab.* Feb;85(2):799-803.

Accepted Manuscript

Pandey, J., Maurya, R., Raykhera, R., Srivastava, M.N., Yadav, P.P & Tamrakar AK. (2014). Murraya koenigii (L.) Spreng. ameliorates insulin resistance in dexamethasone-treated mice by enhancing peripheral insulin sensitivity. *J Sci Food Agric*. 94(11):2282-8.

Pasieka, A. M., & Rafacho, A. (2016). Impact of Glucocorticoid Excess on Glucose Tolerance: Clinical and Preclinical Evidence. Metabolites, 6(3), 24.

Pazini, F. L., Cunha, M. P, Azevedo, D., Rosa, J. M., Colla, A., de Oliveira, J., Ramos-Hryb, A. B., Brocardo, P. S., Gil-Mohapel, J., & Rodrigues, A. L. S. (2017) Creatine Prevents Corticosterone-Induced Reduction in Hippocampal Proliferation and Differentiation: Possible Implication for Its Antidepressant Effect. *Mol Neurobiol*, 54(8):6245-6260.

Poggioli, R., Ueta, C. B., Drigo, R. A., Castillo, M., Fonseca, T. L., & Bianco, A. C. (2013). Dexamethasone reduces energy expenditure and increases susceptibility to diet-induced obesity in mice. *Obesity (Silver Spring, Md.)*, *21*(9), E415–E420.

Protzek, A. O., Costa-Júnior, J. M., Rezende, L. F., Santos, G. J., Araújo, T. G., Vettorazzi, J. F., Boschero, A. C. (2014). Augmented β-Cell Function and Mass in Glucocorticoid-Treated Rodents Are Associated with Increased Islet Ir-β /AKT/mTOR and Decreased AMPK/ACC and AS160 Signaling. *International journal of endocrinology*, 983453.

Prut, L., & Belzung, C. (2003). The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: a review. *Eur J Pharmacol*, 28;463(1-3):3-33.

Rafacho, A., Abrantes, J. L., Ribeiro, D. L., Paula, F. M., Pinto, M. E., Boschero, A. C., & Bosqueiro J.R. (2011). Morphofunctional alterations in endocrine pancreas of short- and long-term dexamethasone-treated rats. *Horm. Metab. Res*, 43:275–281.

Rafacho, A., Cestari, T. M., Taboga, S. R., Boschero, A. C., & Bosqueiro J. R. (2009). High doses of dexamethasone induce increased beta-cell proliferation in pancreatic rat islets. *Am. J. Physiol. Endocrinol. Metab*. 296: E681–E689.

Rafacho, A., Marroquí, L., Taboga, S. R., Abrantes, J. L., Silveira, L. R., Boschero, A. C., Carneiro, E. M., Bosqueiro, J. R., Nadal, A., & Quesada I. (2010). Glucocorticoids in vivo induce both insulin hypersecretion and enhanced glucose sensitivity of stimulus-secretion coupling in isolated rat islets. *Endocrinology*, 151:85–95.

Ramos, A. (2008). Animal models of anxiety: do I need multiple tests? *Trends Pharmacol. Sci.*, 29: 493–498.

Skupio, U., Tertil, M., Sikora, M., Golda, S., Wawrzczak-Bargiela, A & Przewlocki, R. (2015). Behavioral and molecular alterations in mice resulting from chronic treatment with dexamethasone: relevance to depression. *Neuroscience*, 12;286:141-50.

Steru, L., Chermat, R., Thierry, B., & Simon, P. (1995). The tail suspension test: a new method for screening antidepressants in mice. *Psychopharmacology (Berl)*, ;85(3):367-70.

Sztainberg, Y., Kuperman, Y., Tsoory, M., Lebow, M., Chen, A. (2010). The anxiolytic effect of environmental enrichment is mediated via amygdalar CRF receptor type 1. *Mol Psychiatry*, 15(9):905-17.

Accepted Manuscript

Tang, F., Wang, G., & Lian, Y. (2017). Association between anxiety and metabolic syndrome: A systematic review and meta-analysis of epidemiological studies. *Psychoneuroendocrinology*. 77:112-121.

Thunhorst, R. L., Beltz, T. G., & Johnson, A. K. (2007). Glucocorticoids increase salt appetite by promoting water and sodium excretion. *American journal of physiology. Regulatory, integrative and comparative physiology*, *293*(3), R1444–R1451.

Tipton, K. D., & Wolfe, R. R. (2001). Exercise, protein metabolism, and muscle growth. *Int J Sport Nutr Exerc Metab*. 11(1):109-32.

Tonolo, G., Fraser, R., Connell, J. M., & Kenyon, C. (1988). Chronic low-dose infusions of dexamethasone in rats: effects on blood pressure, body weight and plasma atrial natriuretic peptide. *J Hypertens*, 6(1):25-31.

van Praag, H., Kempermann, G., & Gage, F. H. (2000). Neural consequences of environmental enrichment. *Nat Rev Neurosci*, 1(3):191-8.

Vancampfort, D., Stubbs, B., Mitchell, A. J., De Hert, M., Wampers, M., Ward, P. B., … Correll, C. U. (2015). Risk of metabolic syndrome and its components in people with schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder: a systematic review and meta-analysis. *World psychiatry: official journal of the World Psychiatric Association (WPA)*, *14*(3), 339–347.

Wróbel, A., Serefko, A., Wlaź, P., & Poleszak, E. (2014) The depressogenic-like effect of acute and chronic treatment with dexamethasone and its influence on the activity of antidepressant drugs in the forced swim test in adult mice. *Prog Neuropsychopharmacol Biol Psychiatry*, 54:243–248.

Wu, Z., & Fang, Y. (2014). Comorbidity of depressive and anxiety disorders: challenges in diagnosis and assessment. *Shanghai archives of psychiatry*, 26(4), 227–231.

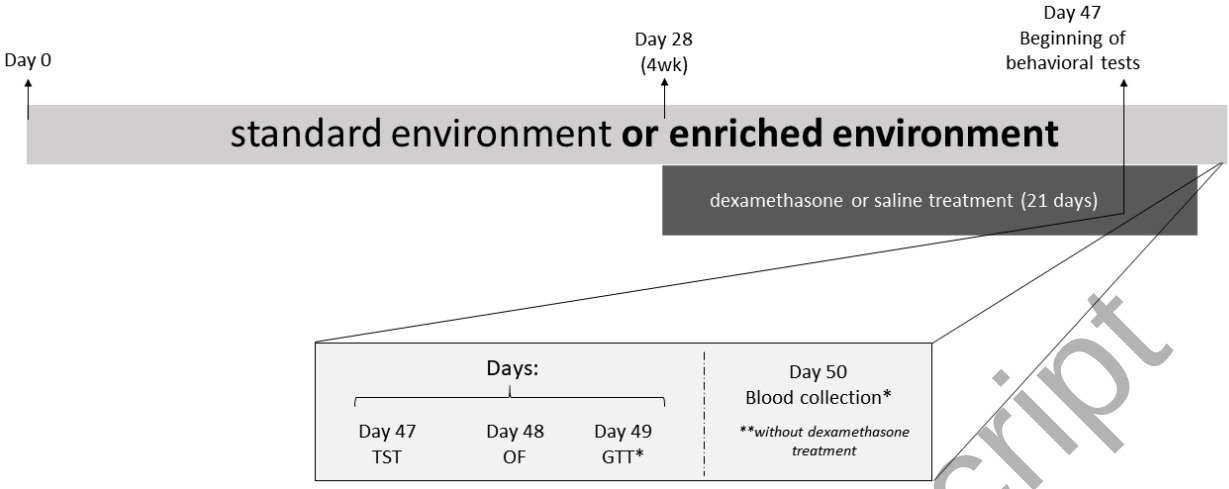
Young, D., Lawlor, P. A., Leone, P., Dragunow, M & During, M. J. (1999) Environmental enrichment inhibits spontaneous apoptosis, prevents seizures and is neuroprotective. *Nat Med*. 1999 Apr;5(4):448-53.

Yu, C. Y., Mayba, O., Lee, J. V., Tran, J., Harris, C., Speed, T. P., & Wang, J. C. (2010). Genome-wide analysis of glucocorticoid receptor binding regions in adipocytes reveal gene network involved in triglyceride homeostasis. PloS one, 5(12), e15188.

Zhang, K., Yang, J., Wang, F., Pan, X., Liu, J., Wang, L., Su, G., Ma, J., Dong, Y., Xiong, Z., & Wu C. (2016) Antidepressant-like effects of Xiaochaihutang in a neuroendocrine mouse model of anxiety/depression. *J Ethnopharmacol*, 24;194:674-683.

Accepted Manuscript

Figure Legends



Legend: TST- tail suspension test; OF- open field; GTT- glucose tolerance test.
**6-h fasting.*

Figure 1. Experimental design.

Accepted Manuscript

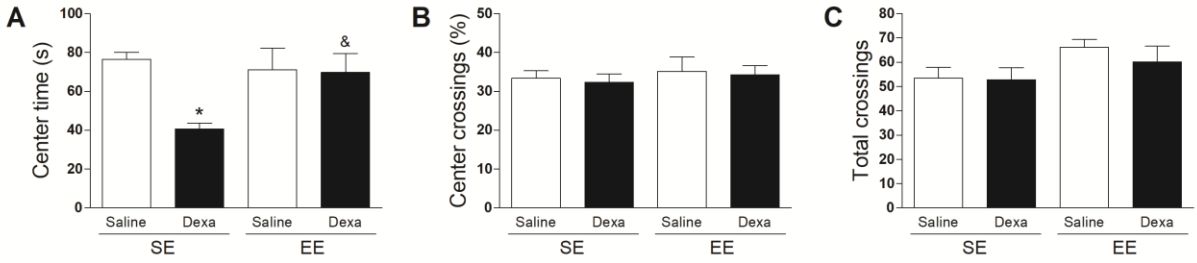


Figure 2. EE mitigates the dexamethasone-induced anxiety-like behavior in the open field test. (A) Center time (s), (B) Percentage of center crossings, (C) Total crossings. Data are expressed as mean + SEM (n = 10 animals per group). *p < 0.05 *vs.* Saline/SE group and &p < 0.05 *vs.* Dexa/SE group (two-way analysis of variance followed by Newman–Keuls *post-hoc* test). Dexa (dexamethasone); SE (standard environment); EE (environmental enrichment).

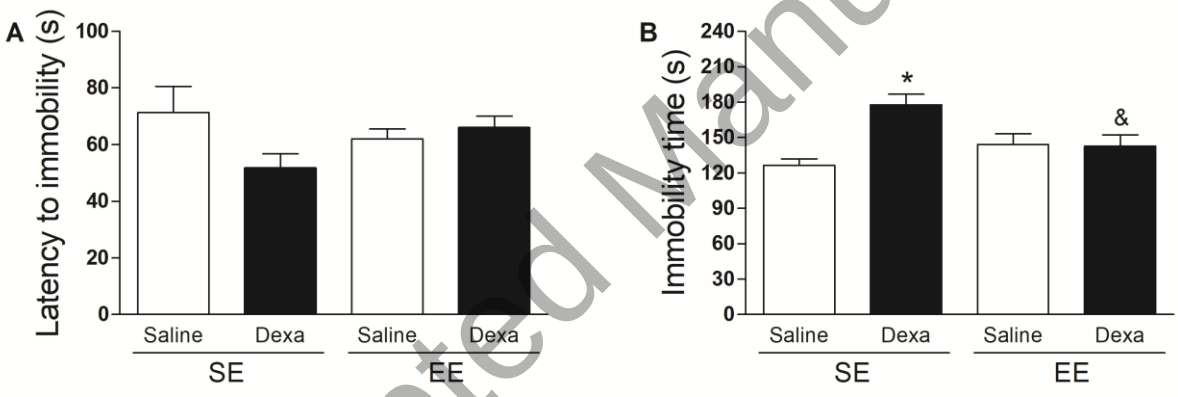


Figure 3. EE mitigates the dexamethasone-induced behavioral despair in the tail suspension test. (A) Latency to immobility (s) and (B) Immobility time (s). Data are expressed as mean + SEM (n = 10 animals per group). *p < 0.05 *vs.* Saline/SE group and &p < 0.05 *vs.* Dexa/SE group (two-way analysis of variance followed by Newman–Keuls *post-hoc* test). Dexa (dexamethasone); SE (standard environment); EE (environmental enrichment).

Accepted Manuscript

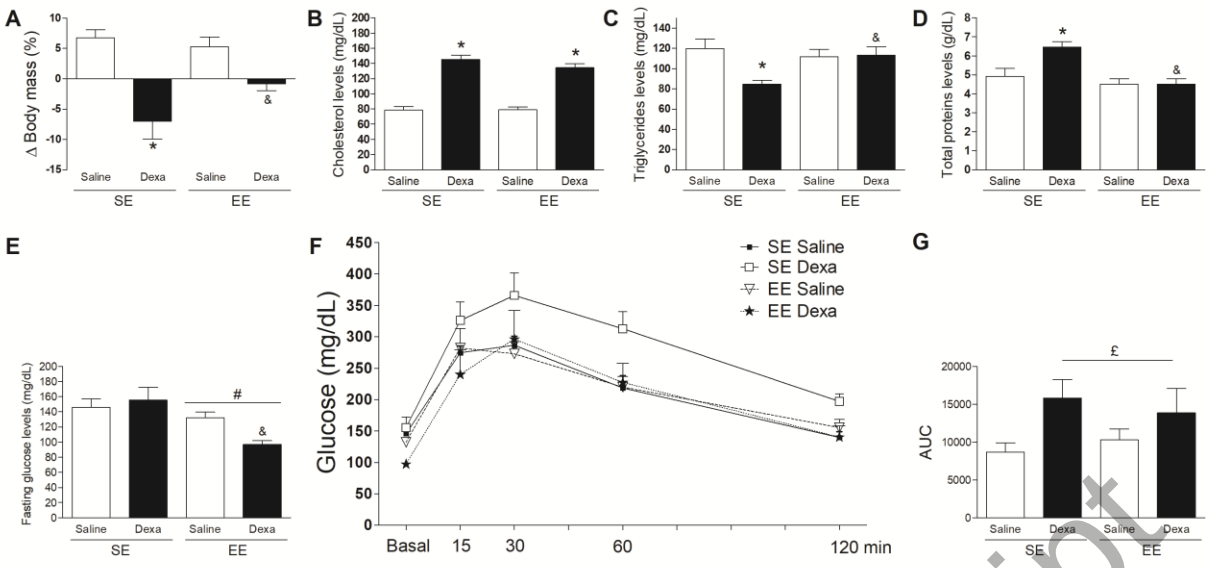


Figure 4. EE effects on dexamethasone-induced metabolic alterarions. (A) Body mass variation (%), (B) Plasma cholesterol levels, (C) Plasma triglycerides levels, (D) Plasma total proteins levels, (E) Basal glucose levels, (F) Glucose tolerance test, and (F) Area under curve of GTT. Data are expressed as mean + SEM (n = 10 animals per group). *p < 0.05 *vs.* Saline/SE group and &p < 0.05 *vs.* Dexa/SE group (two-way analysis of variance followed by Newman–Keuls *post-hoc* test). £p < 0.05 treatment factor (two-way analysis of variance followed). #p < 0.05 environment factor (two-way analysis of variance). Dexa (dexamethasone); SE (standard environment); EE (environmental enrichment).

Accepted Manuscript

Molecular Psychiatry (2000) 5, 467–475
© 2000 Macmillan Publishers Ltd All rights reserved 1359-4184/00 $15.00
www.nature.com/mp

MILLENNIUM ARTICLE

Is there an evolutionary mismatch between the normal physiology of the human dopaminergic system and current environmental conditions in industrialized countries?

L Pani

CNR Center for Neuropharmacology, 'BB Brodie' Department of Neuroscience, University of Cagliari and Neuroscienze Scarl, Cagliari, Italy

A large body of evidence has recently defined a field theory known as 'evolutionary mismatch', which derives its attributes largely from the fact that current environmental conditions are completely different from those in which the human central nervous system evolved. Current views on the evolutionary mismatch theory lack, however, any attempts to define which brain areas or neuronal circuits should be mostly involved in coding such misevolved traits and to what extent our neurobiological knowledge can be applied to the topographical localization of a specific psychopathology. In this respect the mesocorticolimbic dopaminergic circuits have long been misconceptualized as simple reward or reinforcement systems. Instead, they motivate and coordinate the functions of the higher brain areas that mediate planning and foresight and direct finalized movement in both animals and humans. These systems make animals intensely interested in exploring the world around them, but by the same means they also make them susceptible to the environmental stimuli that have been sought and consumed. It is has been speculated that the cortical dopamine targets that developed most recently in phylogeny are of particular functional value, and that the mesocorticolimbic dopaminergic system is involved in more complex integrative functions than previously assumed. In the present paper I will argue that some mental disorders may have their deep roots in the evolutionary mismatch between the normal physiology of the mesocorticolimbic dopaminergic system and the current environmental conditions in affluent societies. *Molecular Psychiatry* (2000) **5,** 467–475.

Keywords: evolution; limbic system; dopamine; stress; depression; emotions; Darwinian medicine

Introduction

The essential steps in the evolution of the brain from early mammals[1] to humans[2] show an increasing motor[3] and cognitive-emotional interaction[4] with the external world (Table 1). This 'environmental pressure' required the development of a neuronal system which was able to decipher and cope with the signals and stimuli coming from the brain's surroundings.

In this sense, the utilization of 'the principle of reward'[5] by linking outcomes of behaviors to either reinforcing or aversing sensations (and memorizing them), which is part of the normal physiology of the dopaminergic system, was absolutely instrumental in the course of evolution.

Several authors, however, consider such a view an over-simplification of the role of the mesolimbic and mesocortical dopamine circuits, which not only direct finalized movement in both animals and humans, but also motivate and coordinate the functions of higher brain areas that mediate planning, foresight and promote states of eagerness.[6,7] By these functions animals become intensely interested in exploring the world around them, but by the same means they are exposed and susceptible to the environmental stimuli that have been sought and consumed. In mammals, and especially in primates, the mesocortical dopaminergic

Table 1 Steps in brain evolution with increasing environmental interaction

1.	Bipedalism
2.	Development of a cooling system
3.	Trichromatic vision
4.	Hemispheric dominance/use of tools
5.	Nursing and maternal care
6	Voluntary control of speech
7.	Separation call
8.	Self/other feelings awareness
9.	Development of synthatic ability
10.	Pressure to speak

Correspondence: Dr L Pani, CNR Center for Neuropharmacology, Via Porcell, 4, 09124-I, Cagliari, Italy. E-mail: panil@unica.it
Received 3 January 2000; revised and accepted 10 April 2000

system generates and sustains curiosity, and it is quite efficient at facilitating learning. It is particularly effective in memorizing information about where material resources are located and the best way to obtain the things needed for survival, including food, water, warmth and sex.[8]

While telencephalic dopaminergic nuclei, such as the nucleus accumbens and the caudate-putamen, have been conserved in their gross and microscopic structure throughout evolution from rodents[9] to humans,[10] the structural organization of the cortical dopaminergic circuits, most strictly involved in human pathologies such as depression and psychosis, shows significant differences between rodents and primates.[11,12]

Recent data show that the major development of the cerebral cortex in primates was accompanied (or produced?) by a similar expansion of the dopaminergic input to all cortical layers, with a specific and laminar redistribution of the dopaminergic terminals.[11] It has been speculated that the cortical dopamine targets that developed most recently in phylogeny are to be considered of particular functional value, and that the mesocorticolimbic dopaminergic system is involved in more complex integrative functions than previously assumed.

In the present paper I will argue that the evolved physiology of the mesocorticolimbic dopaminergic system may be mismatched with respect to current environmental conditions and may contribute to the precipitation of human psychiatric disorders.

Evolution of the dopaminergic system

Enzymes

The acquisition of the behaviors modulated by dopaminergic innervation has been so indispensable in survival that the enzymatic machinery needed to synthesize dopamine appeared very early in phylogenesis. One of the most rudimentary nervous systems, that of the sea pansy *Renilla koellikeri (Cnidaria)*, is able to synthesize dopamine; such synthesis is inhibited by alpha-methyl-p-tyrosine.[13] In this invertebrate, the role of dopamine is that of giving bioluminescence and producing muscular contractions, which are most likely to be used to capture food or avoid danger; insufficient to result in such superior animal characteristics as the ability to feel pleasure and pain. The neurochemical basis of the effects of dopamino-mimetic drugs in other distant organisms has not been fully elucidated, but the hyperkinesia induced by cocaine, a drug which blocks the dopamine transporter, in Planaria specimens suggests that the selective site for pre-synaptic dopamine re-uptake is already present in this flatworm.[14] Goldfish show place preference for the water bowls where they have received amphetamine, a drug of abuse that increases dopaminergic transmission.[15] In the cephalochordate amphioxus, a sister group to vertebrates, the presence of dopamine (but not of noradrenaline) has been assayed. In contrast, two distinct genes homologous to the jawed vertebrate dopamine D1 and beta adrenergic receptor genes were shown to be extant in representatives of the earliest craniates, lamprey and hagfish, paralleling a high dopamine and noradrenaline content throughout their brain.[16]

Receptor family

The evolving vertebrate nervous system was always accompanied by major gene duplication events which generated novel organs along with a definite sympathetic system which was needed to monitor and control them. Vertebrate neural pathways synthesizing monoamine neurotransmitters (mostly dopamine and noradrenaline) were subsequently recruited in order to process increased information demands by mediating psychomotor functions, such as selective attention/predictive reward and emotional drive, via the activation of multiple G-protein linked catecholamine receptor subtypes.

At least 1000 million years ago (MYA), all the classical transmitter ligand molecules evolved the ability to activate a wide range of ion channels, resulting in excitation, inhibition and biphasic or multiphasic responses. In this respect, the invertebrate receptors which have so far been cloned show striking homologies with mammalian receptors, indicating that many of the basic receptor subtypes evolved at an early period, probably around 800 MYA.[17]

The evolution of these receptor-mediated events was similarly driven by forces of gene duplication, at the cephalochordate/vertebrate transition. In amphioxus, a single catecholamine receptor gene was found which, based on molecular phylogeny and functional analysis, forms a monophyletic group with both vertebrate dopamine D1 and beta adrenergic receptor classes.[18] These data suggest that a dopamine D1/beta receptor gene duplication was required for the elaboration of novel catecholamine psychomotor adaptive responses, and that a noradrenergic system specifically emerged at the origin of vertebrate evolution. Recent investigations have confirmed that the dopamine D1A, D1B, and D1C receptors share molecular, pharmacological, and functional attributes that unambiguously allow for their evolutionary classification as distinct D1 receptor subtypes of the G-protein linked super-family, in the vertebrate phylum.[18]

All these receptors conserve a fairly stable secondary structure, a moderate and reasonably steady rate of sequence change, and usually lack introns within their coding sequence. Several results indicate that the first event to occur within this gene family was the divergence of the catecholamine receptors from the muscarinic acetylcholine receptors, which occurred prior to the divergence of the arthropod and vertebrate lineages. Subsequently, the ability to activate specific second-messenger pathways diverged independently in both the muscarinic and the catecholamine receptors. This appears to have occurred after the divergence of the arthropod and vertebrate lineages, but before the divergence of the avian and mammalian lineages. However, the second-messenger pathways activated by adrenergic and dopamine receptors did not diverge independently. Rather, the ability of the catecholamine

receptors to bind to specific ligands, such as epinephrine, norepinephrine, dopamine, or octopamine, was repeatedly modified in evolutionary history, and in some cases was modified after the divergence of the second-messenger pathways.[19]

When the distribution of the dopamine D1 and D2 receptors in the brain of a mouse, rat, or cat is compared with that of a monkey, similarities are found only in the basal ganglia; the highest density being present in the caudate-putamen, in the nucleus accumbens, in the olfactory tubercle, and in the substantia nigra. Apart from these nuclei, and particularly in more recent structures such as the cerebellum and the frontal cortex, the absolute and relative values of the dopamine receptor subtypes is completely different in primates, suggesting that the transcriptional control of the gene encoding for the dopaminergic receptors has also undergone profound changes throughout evolution.[20]

The dopaminergic system and the mismatch theory

Large sets of data (for a review see Nesse and Williams)[21] have recently allowed the definition of a field theory called 'evolutionary mismatch', which derives its attributes largely from the fact that current environmental conditions in industrialized countries are completely different from those in which the human central nervous system evolved. The mismatch between evolved traits and contemporary lifestyle, the presence of evolutionary trade-offs (eg selection of some traits over others), and historical constraints (eg how a trait has evolved), has resulted in the poor specialization of selected traits, as well as a particular susceptibility to mental disorders.

To explain the somatic correlates of such an evolutionary view of diseases, it is assumed that pathogens that attack humans evolve faster than we do. If the ever-mutating environmental conditions, where change is mostly due to fast technological advancement (Table 2), could now be assimilated to organic pathogens, their rapid rate of evolution and change would impose a pressure that is not paralleled by an evolution of the brain structures which are able to analyze, process, and elaborate on such rapid variations. For example, a poor fit between strongly evolved traits (eg dependency) and the current urban social environment may account for the increase in frequency of anxiety and depressive disorders in modern societies.[22]

Current views on the evolutionary mismatch theory lack, however, a definition of which brain areas or neuronal circuits should be involved in coding such misevolved traits, and to what extent our neurobiological knowledge can be applied to the topographical localization of a specific psychopathology.

The appetitive and consummatory phases of behaviors triggered by environmental cues which are associated with survival properties (whether rewarding or aversive for the individual) are known to be mediated by sensory-specific dopaminergic neuronal circuits. These dopaminergic systems are involved in those mechanisms which determine satiety to food,[23] sex,[24] and drugs of abuse.[25] In all these conditions, the most important dopamine systems involved are that of the terminal fields of the prefrontal cortex,[26] and that of the nucleus accumbens,[27] both originating from neurons located in the ventral tegmental area of the midbrain. Whether the stimulation of these neurons mediates the incentive salience (reviewed in Berridge and Robinson),[28] or the associative learning[29] capacities of an event, is still a matter of intense debate. According to the first hypothesis, rewarding (and thus also averse) stimuli that have acquired motivational significance (salience) are part of a phenomenon which is neither unitary nor subjective; dopamine-related neural circuits mediating only a specific component of reward, ie the attribution of incentive salience to an otherwise neutral event. In contrast, others have argued that mesolimbic dopaminergic pathways do not code for such an ability, but serve only to respond to specific stimuli which possess high motivational impact, as a result of their novelty, unpredictability, specific sensory modalities, or occurrence under a deprivation state;[29] in these conditions, dopamine neurons define the substrate for precise associative learning abilities.

From an evolutionary perspective, the two mechanisms must play an equally important role in the attribution of an incentive value to ethologically relevant stimuli, and must facilitate, by means of both incentive-salience and associative-learning, the selection and the memory of appropriate behavioral responses. The conservation of such principles throughout evolution explains why pharmacological manipulations of the dopaminergic transmission are similarly able to modify basic behaviors in distant species such as lizards and rats (Table 3). Stimulation of the dopaminergic transmission will increase these behaviors, while the blocking of dopaminergic receptors or lesions of the dopaminergic pathways will decrease them. This confirms that dopamine is an

Table 2 Years of significant inventions and technological advancement

1550	Screwdriver
1690	Steam engine
1770	Electrical battery
1853	Hypodermic needle
1876	Telephone
1885	Gas-engine auto
1886	Coca-Cola
1903	Airplane
1920	Radio broadcast
1926	Television
1948	Transistors
1962	Minicomputer
1969	Moon landing
1972	Videogames
1978	Test tube baby
1988	Patented animal life
1992	Global internet
2000	Human genome sequenced

Table 3 Special forms of basic behaviors[a]

1.	Establishing a 'territory'
2.	Showing place preference
3.	Hunting
4.	Greeting
5.	Social grouping
6.	Feeding and drinking
7.	Grooming
8.	Courtship
9.	Mating
10.	Breeding
11.	Maternal behavior
12.	Play (mammals only)

[a]Adapted from Reference 6.

important neurotransmitter coding for sensory-specific responses, in the sense that a perceived environmental clue is acted on cognitively, either with an increase in exploration (by means of curiosity and reward) or a decrease (through fear and aversion),[30] but also show that the physiological responses of the dopaminergic system are strictly 'environment sensitive'. Electrophysiological recordings have indicated that key cortical and subcortical dopaminergic regions are involved in these motivated behaviors. Among these, a particular role in primates is given to the connections between the nucleus accumbens and the frontal cortex,[31] which I will consider as primary sites for the evolutionary mismatch between the dopaminergic system and the ability to process significant environmental-driven change-related information that guides coherent behavioral responses.

Expansion and regression of dopaminergic innervation

Animals with big brains are scarce, because bigger neuronal structures are more costly in terms of energy expenditure, development time, and anatomical complexity. It was always supposed that a much better diet (in terms of caloric value) must have been the starting point for the development of a much larger cerebral mass. Recently it was described how, about 40 MYA, a duplication of the gene coding for a retinal cone pigment in an ancestor of prosimians resulted in the development of trichromatic vision.[32] This ability to see in color was instrumental in detecting ripe fruit, and gave an immediate evolutionary advantage over green-leaf eaters. At about the same time, the system for emotional communication via facial expression expanded and the olfactory bulb completed its regression in primates.

Recently, this hypothesis was confirmed, showing that a highly predictable relationship exists between the relative size of the neocortex and that of the total brain.[33] When the sizes of 10 measured brain divisions from 131 species are plotted as a function of total brain size, the two most expanded structures in all species are, in fact, the neocortex and the striatum (caudate-putamen; nucleus accumbens and internal capsule), at variance with that of the olfactory bulb. Thus, two heavily innervated dopaminergic structures—one coupled to taste related reward and satiety, and the other to olfactory satiety—seemed to be mutually exclusive in their development. These complementary characteristics have been conserved in modern humans.[34] The neural mechanisms of sensory-specific satiety in primates are implemented in the orbitofrontal cortex, which has an important integrative function. Neurons in this region of the cortex were indeed found to respond to stimulation of the taste, olfactory, or visual systems. In addition, some neurons were found to possess a bimodal response; responding, for example, to both taste and olfactory, or taste and visual stimuli. Since these multimodal neurons were found in very close proximity to unimodal neurons, and since the unimodal sensory neurons were intermingled, it is possible that the orbitofrontal cortex represents the first cortical area of convergence for these three modalities in primates.[35]

Conditions that have changed through evolution and which interfere with the function of the dopaminergic system

Empirical data suggest a causal relationship between evolutionary-new environmental factors, specific dysregulation of the physiological ability of the dopaminergic system to process environmental stimuli and specific psychiatric disorders. Accordingly, the use of psychotropic medication in outpatient medical practice rose dramatically during the last 10 years, especially in industrialized countries.

Around 10 years ago, Gershon *et al* showed that the rates of bipolar, schizoaffective, and unipolar disorders were higher in the cohorts born after 1940 than in the cohorts born earlier.[36] It was suggested that an ominous trend could be present, leading to an increase in prevalence of a broad spectrum of familial affective disorders in the coming decades. This prediction proved to be correct and confirmed by more recent reports where changes in the prescribing patterns of psychotropic medications by office-based United-States physicians were examined.[37] In the 9 years period between 1985 and 1994, the number of visits during which a psychotropic medication was prescribed increased from 32.73 million to 45.64 million. Antianxiety or hypnotic drug visits, previously the largest category, decreased and were surpassed by antidepressant visits, which doubled from 10 to 20 million in the last 5 years period; and stimulant drug visits increased more than fivefold in the same period. These dramatic trends are in agreement with our findings (Table 4), where we consider the Italian sales of psychotropic medications classified into three major categories: anxiolytics, antidepressants, and antipsychotics and confirmed a significant tendency toward increase for antidepressants (+34%) and antipsychotics (+15%) but not for anxiolytics (+4%) in the last 5 years.

Thus, it appears that in current affluent societies the

Table 4 The Italian market for psychotropic compounds

	Millions of prescriptions per year				
	1995	*1996*	*1997*	*1998*	*1999*
Antipsychotics	11675	12432	12468	12940	13428
Antidepressants	16750	18275	19413	20524	22493
Anxiolytics	33143	34589	34093	34917	34487

frequency of psychiatric disorders requiring medications which modulate the function of (and not only of) the dopaminergic system is increasing. In addition, the outcomes of acute affective disorders are considerably more favorable in developing countries than in comparable studies in developed settings[38] where at least four areas of evolutionary novelty can contribute and will be briefly discussed below.

Drugs of abuse

The ability to extract, purify and lately synthesize psychoactive compounds has increased their availability to the general population in quantities and qualities never experienced before in the history of mankind. In addition, the development of new routes of administration (eg hypodermic needles, crack smoking, organic solvents) has increased the penetration of drugs of abuse into the brain, and has contributed to a change in the traditional relationship between the use of the natural source of the drug (eg the chewing of cocaleaves) to that of its active component (eg the endovenous injection of cocaine hydrochloride).

Drugs of abuse offer a remarkable example of evolutionary mismatch at the level of the dopaminergic system, and help to clarify its role in the general physiology of the central nervous system. If drugs of abuse that act on the dopaminergic system (that is all of them, directly or indirectly) were used solely for their hedonic and 'reward producing' potential, then this capacity, ie the pure ability to experience pleasure, would be the only one to be mainly affected by their chronic use. Instead, after a variable period of use, all drugs which determine a state of dependence interfere with the global adaptation of an individual to its environment, producing not only an impairment of his/her hedonic capacities, but also a more disruptive effect on the cognitive and emotional abilities that are necessary for an effective interaction with the external world (maladaptation). The limbic portion of the dopaminergic system is essential for the generation of the basic emotions that mediate various pro-social behaviors, such as maternal care, playfulness, and friendship, and which have been learned and have evolved to influence motivational states and, ultimately, to increase fitness. All these behaviors, along with the emotions which are related to them and their cognitive implications, are profoundly altered in a drug addict, much more than just the capacity to experience pleasure or avoid pain (whether physical or psychological). Drugs of abuse that convey a signal that falsely indicates the arrival of a huge fitness benefit (much higher than food, water, sex, etc), actively change behavioral hierarchies. As a consequence, drug-seeking increases in frequency to the point where it becomes the only dominant behavior, displacing all the other adaptive traits.[39]

Stress

The definition of the term stress has gone from 'a nonspecific response of the body to any demands made upon it',[40] to 'both a survival mechanism and an indicator of internal and external cues'.[41] Both acute and chronic stress have a detrimental impact on the normal function of the dopaminergic system (for a recent review see Pani *et al*).[42] The response of the dopaminergic system to stress is neither homogenous nor generalized, and numerous results suggest that mesoprefrontal cortex dopaminergic projections are selectively activated by sudden changes in environmental conditions, whether expected or not.[43]

The fine tuning of dopamine extracellular concentration in the medial prefrontal cortex seems to be essential for the decision-taking protocols of the medial prefrontal cortex in the rat and the monkey.[44] This supports the idea of a critical range of dopamine turnover for optimal prefrontal cortical cognitive functioning, with reduced or excessive dopamine turnover leading to cognitive impairment. Several data point to ventral tegmental area projections,[45] dopamine receptors,[46] and a loss of inhibitory tone on ventral tegmental area dopamine cells[47] and lateral-basolateral amygdala (see below) as important regulatory sites for maintaining superior cognitive functions. The increased ability to remove dopamine in chronically stressed animals has also indicated that altered dopamine clearance may serve as an adaptive mechanism in the medial prefrontal cortex.[48] It has been further suggested that increased dopamine D1 receptor stimulation during stress may serve to take the medial prefrontal cortex 'off-line', in order to allow posterior cortical and subcortical structures to regulate more 'primitive' forms of behavior.[49] Since dopaminergic innervation of the medial prefrontal cortex is able to regulate the activity of subcortical dopamine innervations, disruption of the medial prefrontal cortex dopamine fibers may result in the altered biochemical responsiveness of the dopamine subcortical innervations.[50]

In humans, the balance between cortical and subcortical dopaminergic activity may serve as a protection against psychotic decompensation from chronic endogenous or exogenous insult,[51] and the failure of this coping mechanism may well contribute to the vulnerability of the medial prefrontal cortex in many neuropsychiatric disorders related to the stress response.[52]

Chronic emotional reactions

Emotional responses can be difficult to differentiate from generic reactions to stress, in laboratory animals as well as in humans. In addition, emotions such as

empathy seem to be uniquely human, and therefore impossible to mimic in animal models. In spite of this, several animal behaviors such as freezing, self-grooming, exploration and defecation have been considered as valid somatic correlates to human emotional responses in which the dopaminergic system is involved. In all the models, the chronic nature of the stimulus is essential to produce a stable change in the normal physiology of the dopaminergic system. In modern times this can be obtained by generic stress due to sleep disruption, time shifts, poor physical activity, increased physical constraints, and unnatural social rules. In diseases in which dopamine function is compromised, humans exhibit a constellation of emotional symptoms, suggesting that the dopamine system plays an important role in the integration of superior cortical functions. A dorsal tier of dopamine neurons receives input from the nucleus accumbens and from the amygdala, and projects widely throughout the cortex. Through these projections, the dopaminergic limbic system has an enormous influence on cortical output and can therefore affect the emotional and motivational 'coloring' of a wide range of behaviors.[53]

In this respect, the lateral-basolateral amygdala receives sensory input from environmental stimuli and, along with the central amygdala, mediates conditional fear. Through its projections to forebrain, midbrain and hindbrain areas, the central amygdala governs the behavioral, autonomic, and endocrine responses that characterize a central fear state.[54] Accumulating revelations about the amygdala-based fear system has led to considerable progress in delineating the neural connections and cellular mechanisms that underlie aversive emotionality, and very recently electrophysiological evidence has shown that neural discharge in the central amygdala is a consequence of repeated low-current, high-frequency electrical stimulation of dopaminergic nuclei of the ventral tegmental area.[55]

The introduction of Roman high-avoidance (RHA) and Roman low-avoidance (RLA) rats, two selectively bred lines that differ in their level of emotionality, into an unfamiliar environment, the application of a high-intensity loud noise, or immobilization, are all associated with an increase in extracellular cortical dopamine metabolite levels in the RHA but not in the RLA line, suggesting that the differently evoked emotional reactions may produce a common activation of cognitive processes.[56] This difference in reaction to a stressful stimulus, according to the genetic makeup of an individual, is an essential function of the mesencephalic dopaminergic neurons, and of their role in the organization of behavioral responses which are the reflection of a heightened attention of the animal attempting to cope with the stressor.[57] The mesocortical dopaminergic innervations seem to facilitate the functions of integrative structures, and to have a coordinated role in regulating the information flow between cortical structures.[58]

The study of fear and the role of the dopaminergic system in rodents have, however, produced inconsistent data. While some results suggest an important relationship between social and motor reactivity, as well as an important, albeit again strain-dependent, role for dopamine D1 receptors in mediating specific emotional behaviors,[59] others demonstrate that long-term administration of low, pre-synaptic doses of the dopamine D2 antagonist l-sulpiride is highly effective in an animal model of anticipatory anxiety/panic behavior.[60] In contrast, recent findings suggest that quinpirole (a dopamine D2 agonist) decreases fear by blocking the retrieval of a learned association between a conditioned and an unconditioned stimulus.[61]

Chronic sleep deprivation

Humans have conserved mechanisms, like those that exist in other animals, which detect changes in day length and make corresponding adjustments in the duration of nocturnal periods. Several studies from Wehr and collaborators[62] have suggested that modern men's use of artificial light after dark and artificial darkness during the daytime suppresses responses[63] to seasonal changes in the duration of the natural scoto- and photoperiod that might otherwise occur at a given latitude, producing a chronic disturbance in the sleep pattern and decreasing, overall, the total amount of sleep. It is difficult to conceive that chronic sleep deprivation in humans could be obtained and maintained without any activation in the stress axis.

Sleep deprivation in rats has been extensively studied as a possible animal model of mania.[64] Interestingly, at the end of the period of sleep deprivation (approximately 72 h), the rat does not fall asleep as soon as it is returned to its home cage, but shows a period of wakefulness of about 30 min, during which the animal presents a cohort of symptoms that appear to mimic those present in idiopathic mania. In particular, during this period the animal displays a high degree of hyperactivity, irritability, aggressiveness, hyper-sexuality, and stereotypy. The model allowed the discovery of an active role of limbic dopamine in the generation of arousal and insomnia related to sleep deprivation-induced stress.[65] Accordingly, direct evidence demonstrates that acute and chronic sleep deprivation may trigger a manic episode in otherwise healthy individuals,[66] and that recurrent brief hypomania which lasts 1–3 days and belongs to the bipolar spectrum, is increasing in young adults.[67] These findings are indirectly confirmed by the increased use of neuroleptics (Table 4) which, although their use has been widely discouraged for mood disorders, are still the standard treatment for acute bipolar mania.[68]

Conclusions

The brain of *Homo sapiens sapiens* developed on the African plains, in populations of a few hundred thousand individuals; today there are six billion of us worldwide.

It is time to ask ourselves whether the structural organization and the neuronal plasticity of the human

brain are still able to control the evolution of the environment he has created. Some influential neurobiologists have expressed strong doubts.[69]

The dopaminergic innervation to the prefrontal cortex is essential for processing and elaborating on the 'expectancy' of events, and in the working memory related to them, in both rats and monkeys.[44] Environmentally induced changes in the activity of the dopaminergic system, either by conventional (food, water, sex etc) or non-conventional (drugs of abuse, chronic stress, emotional reactions etc) stimuli could mismatch the natural codes for reward-aversion, sensory-motor, incentive-salience mechanisms which have been conserved in the last millions of years and which are needed to associate a learned outcome of a given stimuli with a definite emotional state of the organism.

In this sense, the dorsolateral portion of the prefrontal cortex is important in monitoring the context of the situation in relation to how, and what, reward is given.[70] Patients with prefrontal lesion are poor in assessing how actions relate to goals. The results suggest that pre-frontal cortical lesions provoke a selective impairment in managerial knowledge that may contribute to difficulties in the formulation and execution of plans.[71]

Neurons which are critical for the computation of sensory-specific satiety in the monkey are located in the orbitofrontal cortex and, in humans, decision-making processes depend on signals arising in the prefrontal cortex in association with peripheral somato-visceral signals that can introduce biases in the reasoning processes.[72] When normal subjects experience situations that they associate with the possibility of negative consequences, they sense a peripheral reaction (mainly an adrenergic response) and learn to make decisions accordingly. Damage to the orbitofrontal cortex in adulthood as well as in early life[73] disrupts the capacity to make these physiological responses, and impairs this ability to make crucial social decisions.

Considering the profound influence of prefrontal cortex damage on several adaptive behaviors in rats and primates, another question arises: is it possible that in the absence of organic pathologies, a 'functional uncoupling' of the cognitive capacities (eg working memory, decision-making, associative learning) of the prefrontal cortex could be the site of the dopaminergic evolutionary mismatch?

In this respect, the temperament (as the genetic model of personality)[74] of an individual may strongly relate to this evolutionary model. If this theory holds true it should be possible to study by imaging techniques whether or not patients with different temperaments, obviously selected by natural selection for conferring evolutionary trade-off, such as the bipolar-anxious temperament types,[75,76] would show any difference in the functions of their prefrontal cortex activity.

A better understanding of the mechanisms, origins and functions of the dopaminergic system as well as other neurotransmitters and modulators, will also enhance our ability to cope with stress-related mental disorders in the modern environment, and sharpen our wisdom in making decisions about the therapeutic use of psychoactive drugs.

Acknowledgements

I am grateful to Anna Porcella for her critical comments and to David Webb for his precise editorial work on the manuscript. This work was supported by a CNR grant.

References

1 Divac I. Monotremunculi and brain evolution. *TINS* 1995; **18**: 2–4.
2 Wilson DR. Evolutionary epidemiology. Darwinian theory in the service of medicine and psychiatry. *Acta Biotheoretica* 1993; **41**: 205–218.
3 Commentary on Greenfield PM, Language tools and brain: the ontogeny and phylogeny of hierarchically organized sequential behavior. *Behav Brain Sci* 1994; **17**: 357–365.
4 Povinelli DJ, Preuss TM. Theory of mind: evolutionary history of a cognitive specialization. *TINS* 1995; **18**: 418–424.
5 Wise RA. The brain and reward. In: Liebman J, Cooper SJ (eds). *The Neuro-pharmacological Basis of Reward*. Oxford University Press: Oxford, 1989, pp 248–262.
6 MacLean PD. Brain evolution relating to family, play and the separation call. *Arch Gen Psychiatry* 1985; **42**: 405–417.
7 Le Moal M, Simon H. Mesocorticolimbic dopaminergic network: functional and regulatory roles. *Physiol Rev* 1991; **71**: 155–234.
8 Panksepp J. *Affective Neuroscience: the Foundations of Human and Animal Emotions*. Oxford Univ Press: 1998, pp 53–55.
9 Bertler A, Rosengren E. Occurrence and distribution of catecholamine in brain. *Acta Physiol Scand* 1959; **47**: 350–361.
10 Sano I, Gamo T, Kakimoto Y, Taniguchi K, Takesada M, Nishinuma K. Distribution of catechol compounds in human brain. *Biochem Biophys Acta* 1959; **32**: 586–588.
11 Berger B. Dopaminergic innervation of the frontal cerebral cortex. Evolutionary trends and functional implications. In: Chauvel P, Delgado-Escueta (eds). *Advances in Neurology*, Vol 57, Raven Press: New York, 1992, pp 525–544.
12 Berger B, Gaspar P, Verney C. Dopaminergic innervation of the cerebral cortex: unexpected differences between rodents and primates [erratum appears in *TINS* 1991; **14**: 119]. *TINS* 1991; **14**: 21–27.
13 Pani AK, Anctil M. Evidence for biosynthesis and catabolism of monoamines in the sea pansy *Renilla koellikeri (Cnidaria)*. *Neurochem Int* 1994; **25**: 465–474.
14 Palladini G, Ruggeri S, Stocchi F, De Pandis MF, Venturini G, Margotta V. A pharmacological study of cocaine activity in planaria. *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol* 1996; **115**: 41–45.
15 Lett BT, Grant VL. The hedonic effects of amphetamine and pentobarbital in goldfish. *Pharmacol Biochem Behav* 1989; **32**: 355–356.
16 Vincent JD, Cardinaud B, Vernier P. Evolution of monoamine receptors and the origin of motivational and emotional systems in vertebrates. *Bull Acad Natl Med* 1998; **182**: 1505–1516.
17 Walker RJ, Brooks HL, Holden-Dye L. Evolution and overview of classical transmitter molecules and their receptors. *Parasitology* 1996; **113**: S3–S33.
18 Cardinaud B, Sugamori KS, Coudouel S, Vincent JD, Niznik HB, Vernier P. Early emergence of three dopamine D1 receptor subtypes in vertebrates. Molecular phylogenetic, pharmacological, and functional criteria defining D1A, D1B, and D1C receptors in European eel *Anguilla anguilla*. *J Biol Chem* 1997; **272**: 2778–2787.
19 Fryxell KJ. The evolutionary divergence of neurotransmitter receptors and second-messenger pathways. *J Mol Evol* 1995; **41**: 85–97.
20 Camps M, Kelly PH, Palacios JM. Autoradiographic localization of dopamine D1 and D2 receptors in the brain of several mammalian species. *J Neural Transm Gent Sect* 1990; **80**: 105–127.
21 Nesse RM, Williams GC. *Why We Get Sick*. Random House: New York, 1994.
22 McGuire MT, Troisi A. *Darwinian Psychiatry*. Oxford University Press: New York, 1998.

23 Ahn S, Phillips AG. Dopaminergic correlates of sensory-specific satiety in the medial prefrontal cortex and nucleus accumbens of the rat. *J Neurosci* 1999; **19**: RC29.
24 Fiorino DF, Coury A, Phillips AG. Dynamic changes in nucleus accumbens dopamine efflux during the Coolidge effect in male rats. *J Neurosci* 1997; **17**: 4849–4855.
25 Fibiger HC. Role of catecholamine transmitters in reward systems: implications for the neurobiology of affect. In: Oreland E (ed). *Brain Reward Systems and Abuse*. New York Press: New York, 1987, pp 61–74.
26 Cenci MA, Kalén P, Mandel RJ, Björklund A. Regional differences in the regulation of dopamine and noradrenaline release in medial frontal cortex, nucleus accumbens and caudate-putamen: a microdialysis study in the rat. *Brain Res* 1992; **581**: 217–228.
27 Phillips AG, Atkinson LJ, Blackburn JR, Blaha CD. Increased extracellular dopamine in the nucleus accumbens of the rat elicited by a conditioned stimulus for food: an electrochemical study. *Can J Physiol Pharmacol* 1993; **71**: 387–393.
28 Berridge KC, Robinson TE. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res Rev* 1998; **28**: 309–369.
29 Bassareo V, Di Chiara G. Differential influence of associative and nonassociative learning mechanisms on the responsiveness of prefrontal and accumbal dopamine transmission to food stimuli in rats fed ad libitum. *J Neurosci* 1997; **17**: 851–861.
30 Blackburn JR, Pfaus JG, Phillips AG. Dopamine functions in appetitive and defensive behaviors. *Prog Neurobiol* 1992; **39**: 247–279.
31 Rolls ET. *The Brain and Emotions*. Oxford University Press: New York, 1999.
32 Jacobs GH, Neitz M, Deegan JF, Neitz J. Trichromatic colour vision in New World monkeys. *Nature* 1996; **382**: 156–158.
33 Finlay BL, Darlington RB. Linked regularities in the development and evolution of mammalian brains. *Science* 1995; **268**: 1578–1584.
34 Rolls ET, Rolls JH. Olfactory sensory-specific satiety in humans. *Physiol & Behav* 1997; **3**: 461–473.
35 Rolls ET, Baylis LL. Gustatory, olfactory, and visual convergence within the primate orbitofrontal cortex. *J Neurosci* 1994; **14**: 5437–5452.
36 Gershon ES, Hamovit JH, Guroff JJ, Nurnberger JI. Birth-cohort changes in manic and depressive disorders in relatives of bipolar and schizoaffective patients. *Arch Gen Psychiatry* 1987; **44**: 314–319.
37 Pincus HA, Tanielian TL, Marcus SC, Olfson M, Zarin DA, Thompson J *et al*. Prescribing trends in psychotropic medications: primary care, psychiatry, and other medical specialties. *JAMA* 1998, **279**: 526–531.
38 Brown AS, Varma VK, Malhotra S, Jiloha RC, Conover SA, Susser ES. Course of acute affective disorders in a developing country setting. *J Nerv Ment Dis*. 1998; **186**: 207–213.
39 Nesse RM, Berridge KC. Psychoactive drugs in evolutionary perspective. *Science* 1997; **278**: 63–66.
40 Selye H. The evolution of the stress concept. *Am Sci* 1973; **61**: 692–696.
41 Akil HA, Morano IM. Stress. In: Kupfer D, Bloom F (eds). *Psychopharmacology the Fourth Generation of Progress*. Raven Press: New York, 1995, pp 773–785.
42 Pani L, Porcella A, Gessa GL. The role of stress in the pathophysiology of the dopaminergic system. *Mol Psychiatry* 2000; **5**: 14–21.
43 Kaneyuki H, Yokoo H, Tsuda A, Yoshida M, Mizuki Y, Yamada M *et al*. Psychological stress increases dopamine turnover selectively in mesoprefrontal dopamine neurons of rats: reversal by diazepam. *Brain Res* 1991; **557**: 154–161.
44 Murphy BL, Arnsten AF, Goldman-Rakic PS, Roth RH. Increased dopamine turnover in the prefrontal cortex impairs spatial working memory performance in rats and monkeys. *Proc Natl Acad Sci USA* 1996; **93**: 1325–1329.
45 Murphy BL, Arnsten AF, Jentsch JD, Roth RH. Dopamine and spatial working memory in rats and monkeys: pharmacological reversal of stress, induced impairment. *J Neurosci* 1996; **16**: 7768–7775.
46 Steketee JD, Kalivas PW. Sensitization to psychostimulants and stress after injection of pertussis toxin into the A10 dopamine region. *J Pharmacol Exp Ther* 1991; **259**: 916–924.
47 Sorg BA, Steketee JD. Mechanisms of cocaine-induced sensitization. *Prog Neuropsychopharmacol Biol Psychiatry* 1992; **16**: 1003–1012.
48 Meiergerd SM, Schenk JO, Sorg BA. Repeated cocaine and stress increase dopamine clearance in the rat medial prefrontal cortex. *Brain Res* 1997; **773**: 203–207.
49 Zahrt J, Taylor JR, Mathew RG, Arnsten AF. Supranormal stimulation of D1 dopamine receptors in the rodent prefrontal cortex impairs spatial working memory performance. *J Neurosci* 1997; **17**: 8528–8535.
50 Deutch AY, Clark WA, Roth RH. Prefrontal cortical dopamine depletion enhances the responsiveness of mesolimbic dopamine neurons to stress. *Brain Res* 1990; **521**: 311–315.
51 Friedhoff AJ, Carr KD, Uysal S, Schweitzer J. Repeated inescapable stress produces a neuroleptic-like effect on the conditioned avoidance response. *Neuropsychopharmacology* 1995; **13**: 129–138.
52 Finlay JM, Zigmond MJ. The effect of stress on central dopaminergic neurons: possible clinical implications. *Neurochem Res* 1997; **22**: 1387–1394.
53 Haber SN, Fudge JL. The interface between dopamine neurons and the amygdala: implications for schizophrenia. *Schizophr Bull* 1997; **23**: 471–482.
54 Davis M. The role of the amygdala in conditioned fear. In: Aggleton JP (ed). *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*. Wiley: New York, 1992, pp 255–305.
55 Gelowitz DL, Kokkinidis L. Enhanced amygdala kindling after electrical stimulation of the ventral tegmental area: implications for fear and anxiety. *J Neurosci* 1999; **19**: RC41.
56 Bertolucci-D'Angio M, Serrano A, Scatton B. Mesocorticolimbic dopaminergic systems and emotional states. *J Neurosci Meth* 1990; **34**: 135–142.
57 D'Angio M, Serrano A, Driscoll P, Scatton B. Stressful environmental stimuli increase extracellular DOPAC levels in the prefrontal cortex of hypoemotional (Roman high-avoidance) but not hyperemotional (Roman low-avoidance) rats. An *in vivo* voltammetric study. *Brain Res* 1988; **451**: 237–247.
58 Louilot A, Taghzouti K, Simon H, Le Moal M. Limbic system, basal ganglia, and dopaminergic neurons. Executive and regulatory neurons and their role in the organization of behavior. *Brain Behav Evol* 1989; **33**: 157–161.
59 Gendreau PL, Gariépy JL, Petitto JM, Lewis MH. D1 dopamine receptor mediation of social and nonsocial emotional reactivity in mice: effects of housing and strain difference in motor activity. *Behav Neurosci* 1997; **111**: 424–434.
60 Cavazzuti E, Bertolini A, Vergoni AV, Arletti R, Poggioli R, Forgione A *et al*. l-Sulpiride, at a low, non-neuroleptic dose, prevents conditioned fear stress-induced freezing behavior in rats. *Psychopharmacology (Berl)* 1999; **143**: 20–23.
61 Nader K, LeDoux J. The dopaminergic modulation of fear: quinpirole impairs the recall of emotional memories in rats. *Behav Neurosci* 1999; **113**: 152–165.
62 Wehr TA, Giesen HA, Moul DE, Turner EH, Schwartz PJ. Suppression of men's responses to seasonal changes in day length by modern artificial lighting. *Am J Physiol* 1995; **269**: R173–R178.
63 Wehr TA. Effect of seasonal changes in daylength on human neuroendocrine function. *Horm Res* 1998; **49**: 118–124.
64 Gessa GL, Pani L, Fadda P, Fratta W. Sleep deprivation in the rat: an animal model of mania. *Eur Neuropsychopharmacol* 1995; **5**: S89–S93.
65 Fadda P, Martellotta MC, Gessa GL, Fratta W. Dopamine and opioids interactions in sleep deprivation. *Prog Neuropsychopharmacol Biol Psychiatry* 1993; **17**: 269–278.
66 Wright JB. Mania following sleep deprivation. *Br J Psychiatry* 1993; **163**: 679–680.
67 Angst J. The emerging epidemiology of hypomania and bipolar II disorder. *J Affect Disord* 1998; **50**: 143–151.
68 Chou JC, Zito JM, Vitrai J, Craig TJ, Allingham BH, Czobor P. Neuroleptics in acute mania: a pharmacoepidemiologic study. *Ann Pharmacother* 1996; **30**: 1396–1398.
69 Changeux JP. *L'Homme Neuronal*. Fayard: Paris, 1983.
70 Watanabe M. Reward expectancy in primate prefrontal neurons. *Nature* 1996; **382**: 629–632.
71 Sirigu A, Zalla T, Pillon B, Grafman J, Agid Y, Dubois B. Selective impairments in managerial knowledge following pre-frontal cortex damage. *Cortex* 1995; **31**: 301–316.

72 Damasio AR. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philos Trans R Soc Lond B Biol Sci* 1996; **351**: 1413–1420.
73 Anderson SW, Bechara A, Damasio H, Tranel D, Damasio AR. Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neurosci* 1999; **11**: 1032–1037.
74 von Zerssen D, Akiskal HS. Personality factors in affective disorders: historical developments and current issues with special reference to the concepts of temperament and character. *J Affect Disord* 1998; **51**: 1–5.
75 Akiskal HS. Toward a definition of generalized anxiety disorder as an anxious temperament type. *Acta Psychiatr Scand Suppl* 1998; **393**: 66–73.
76 Perugi G, Toni C, Akiskal HS. Anxious-bipolar comorbidity. Diagnostic and treatment challenges. *Psychiatr Clin North Am* 1999; **22**: 565–583.

Human Nature
https://doi.org/10.1007/s12110-018-9327-1

Eyes, More Than Other Facial Features, Enhance Real-World Donation Behavior

Caroline Kelsey[1] · Amrisha Vaish[1] · Tobias Grossmann[1]

Published online: 15 October 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Humans often behave more prosocially when being observed in person and even in response to subtle eye cues, purportedly to manage their reputation. Previous research on this phenomenon has employed the "watching eyes paradigm," in which adults displayed greater prosocial behavior in the presence of images of eyes versus inanimate objects. However, the robustness of the effect of eyes on prosocial behavior has recently been called into question. Therefore, the first goal of the present study was to attempt to replicate this effect. Additionally, it remains unclear whether the watching-eyes effect is driven specifically by reputation management (owing to the monitoring function of the eyes) or whether any cues indexing human presence more generally also have a similar effect. To address these questions, the current study compared prosocial behavior in the presence of eyes versus inanimate objects as well as other human features. The study was conducted as a field experiment at a children's museum. Each week, the donation signs were changed to show eyes, noses, mouths, or chairs. Total donation amount and number of patrons per week were recorded. Participants donated more when they were exposed to eyes than to inanimate objects (chairs). We thus replicated the previously reported watching-eyes effect. Moreover, more money was donated when individuals were exposed to eyes than to more general cues of human presence (nose and mouth). The current findings suggest that eyes play a special role in promoting cooperation in humans, likely by serving as cues of monitoring and thus eliciting reputation management behavior.

Keywords Cooperation · Prosocial behavior · Reputation · Watching-eyes effect

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12110-018-9327-1) contains supplementary material, which is available to authorized users.

✉ Amrisha Vaish
vaish@virginia.edu

[1] Department of Psychology, University of Virginia, Charlottesville, VA 22904, USA

Humans are highly cooperative beings who willingly offer resources to non-kin on a daily basis (Hare 2017). This high level of generosity is maintained in part by reputational forces—through both the gains incurred when individuals are seen carrying out positive actions and the costs incurred when individuals are seen breaking cooperative norms (Fehr and Gächter 2002; Milinski et al. 2002; Wedekind and Milinski 2000). As a result of these reputational forces, people often engage in reputation management (or "self-presentation"; Goffman 1959) by behaving more prosocially when being watched by others (e.g., Bull and Gibson-Robinson 1981; Kurzban 2001). Indeed, even children as young as 5 years of age are more generous and less antisocial in the presence of others (Engelmann et al. 2012; Leimgruber et al. 2012).

It is argued that reputation management does not rely solely on explicit or conscious reasoning; rather, because managing one's reputation has been so vital for the evolution of human cooperation, natural selection has shaped human psychology to also be sensitive to even minimal cues of being observed (Haley and Fessler 2005). This "watching-eyes" effect is supported by numerous studies. For example, several studies using one-shot anonymous dictator games found that donations (either the proportion of individuals who donated, the average donation amount, or both) were greater from participants exposed to images of eyes than to non-social stimuli (e.g., geometric shapes; Haley and Fessler 2005; Keller and Pfattheicher 2011; Nettle et al. 2013; Oda et al. 2011). In field experiments, the presence of eyes compared with non-social controls has been found, for instance, to increase removal of litter, increase voter turnout, and reduce bicycle theft (Bateson et al. 2013; Burnham and Hare 2007; Ekström 2012; Ernest-Jones et al. 2011; Nettle et al. 2012; Panagopoulos 2014a, b).

However, several studies—from both the laboratory and the field—have failed to find support for the watching-eyes effect (Carbon and Hesslinger 2011; Fehr and Schneider 2010; Manesi and Pollet 2017; Saunders et al. 2016; Sparks and Barclay 2015). In addition, the effect is often apparent only under certain conditions and using certain measures. For instance, one meta-analysis of 25 studies found that increased generosity emerges after short exposure to images of eyes (i.e., when eye cues are made suddenly visible or attention is drawn to them in the seconds before the dependent measure is recorded), but not after long exposures (i.e., when eye cues are available for several minutes prior to when the dependent measure is recorded; Sparks and Barclay 2013). Note, however, that this meta-analysis used a vote-counting technique: It included any study that found either an increase in the proportion of individuals who donated or an increase in donation amount after exposure to eye cues as providing support for the effect. Specifically, studies that reported a significant effect with one measure, but not another measure, were nonetheless included as a vote for the effect, which may have made the effect appear more robust than it is. Another meta-analysis (Nettle et al. 2013) suggested that the watching-eyes effect is not generally apparent in the amount that people donate but rather whether people donate anything at all, and the most recent meta-analysis revealed no evidence that cues of observability increase the proportion of individuals who engage in prosocial behavior or the total amount of prosocial behavior (Northover et al. 2017). Given the mixed evidence on the watching-eyes effect, and given the importance of replicability of experimental work (Nosek et al. 2015), the first goal of the present study was to attempt to replicate the watching-eyes effect. Toward this end, we compared people's generosity in the presence of an image of eyes versus an image of a nonsocial stimulus (a chair).

Importantly, even when the watching-eyes effect does emerge, it remains unclear whether it is indeed driven by reputation management. Typically, studies on this topic have compared the effect of eyes (or eye-like images) with that of images of non-social objects (e.g., flowers or geometric shapes). In these comparisons, greater prosociality in the presence of eyes is taken as evidence that individuals exposed to the eyes feel that they are being observed and thus act in ways that will maintain or enhance their reputations (Bateson et al. 2006; Haley and Fessler 2005). It is thus assumed that it is the monitoring (or reputation-forming) function of eyes that elicits greater prosociality. An alternative possibility, however, is that eyes serve as a more general cue to human presence, and that other general cues of human presence (such as other facial features) could also increase people's prosocial behavior through various possible means, such as by making them feel more social, reminding them that they are part of a group, reminding them that there are people in need, alerting them to the possibility of being punished or sanctioned, and so on. In other words, the mechanism underlying the watching-eyes effect might not be reputation management but rather more general effects of social presence. If this latter argument is correct, then the eyes may not be unique in increasing generosity; rather, any cue of human presence may result in this effect. In contrast, if the watching-eyes effect is at least partially explained by reputation management, then eyes should be especially effective in increasing prosocial behavior, as the eyes are unique among the human features in their "watching" function and are therefore an especially relevant cue for reputation management (see Manesi et al. 2016; Vaish et al. 2017). The second goal of this study was thus to test whether eye cues in particular or cues of human presence more generally enhance prosocial behavior.

Although no prior work has directly tested this question, a few studies have attempted to clarify the mechanisms underpinning the watching-eyes effect. Eyes are highly salient cues, and from birth, humans preferentially attend to eyes (for a review, see Grossmann 2017). This privileged attention to eye cues is seen in newborns' preferences for eye-like stimuli (e.g., configurations of three dots that resemble a face with eyes) and in newborns' preferences for faces with direct gaze as opposed to averted gaze (Farroni et al. 2002, 2005). This preferential sensitivity to eyes continues to exist into adulthood, as adults prefer to look at eyes more than at other body parts (Vaish et al. 2017). Eyes, and especially watching eyes, are thus highly salient cues.

Regarding the watching effect in particular, of most relevance are three experiments in which Manesi and colleagues (Manesi et al. 2016) examined whether adults exposed to "watching" eyes (i.e., eyes with direct gaze) were more likely to help others with a dull cognitive task than individuals exposed to closed or averted eyes and to a non-social control (flowers). The findings across the three experiments were inconsistent: There was evidence for increased helping after exposure to watching eyes under certain conditions (e.g., for open eyes compared with closed eyes, and direct gaze compared with a head turned away) but not under other conditions (e.g., direct gaze versus averted gaze without a head turn). When the results from the three experiments were analyzed together, there was some evidence that participants were more generous in the presence of watching eyes than the other stimuli. Manesi et al. (2016) tentatively concluded that it is the monitoring function of the eyes that increases prosocial behavior and that the mechanism behind the watching-eyes effect is thus reputation management. In line with this conclusion, other work shows that the impact of eye cues on donation behavior is mediated by how much individuals expect their donation behavior

to be evaluated by others and by the expectation that donating more money will enhance their reputation (Oda et al. 2011). Additionally, studies on individual differences show that individuals with higher chronic public self-awareness (who are presumably more concerned about their reputations) are more generous in the presence of watching eyes than individuals who are lower on this trait (Keller and Pfattheicher 2011), and individuals who attend more to eyes (but not to other human features) are more generous, but only when donating publicly (i.e., when their reputation is at stake), not when donating privately (Vaish et al. 2017). This body of work provides some support for the reputation management explanation.

Moreover, recent work has examined how donation amounts given in a dictator game differ after exposure to eyes when compared with other social cues. One study reported that individuals exposed to eyes (photos portraying only the eye regions of statues) donated significantly more money than individuals in a more general social condition (pictures of peers not directly facing the participant) or in a non-social control condition (pictures of school hallways; Baillon et al. 2013). Similarly, Haley and Fessler (2005) examined whether another sign of human presence—auditory cues—influenced donation behavior. They found that adults who wore noise-reducing headphones (and thus received fewer auditory cues of human presence) donated just as much as those who did not wear headphones, suggesting that eyes are perhaps especially effective at eliciting reputational concerns.

Nonetheless, given the mixed findings from Manesi et al.'s (2016) experiments and the fact that no prior work has assessed the effect of other human features on generosity in a real-world setting, we aimed here to further investigate the reputation management explanation of the watching-eyes effect using a novel approach. Specifically, we compared donation behavior in the presence of eyes to that of other human features in order to directly test whether eyes play a special role in enhancing prosocial behavior through eliciting reputational concern. We reasoned that if the presence of eyes increases prosocial behavior by eliciting reputational concerns, then the eyes should be especially effective in increasing prosocial behavior as compared with other human features. On the other hand, if eyes increase prosocial behavior merely by cueing human presence, then other human features should be similarly effective in increasing prosocial behavior. The two other human features we used were the nose and mouth. Like the eyes, they are inner features of the human face and are thus similar to the eyes in terms of their visibility to others but do not serve the visual monitoring function that eyes do.

Finally, to our knowledge, the prior studies using social control conditions were all conducted in laboratory settings (Baillon et al. 2013; Haley and Fessler 2005; Manesi et al. 2016); to examine the generalizability of the effects to a real-world setting, it is important to extend this line of research to the field. The present study was thus conducted as a field experiment. Specifically, the study was conducted in a children's museum. Each week, the donation signs were changed to feature images of eyes, noses, mouths, or chairs, and the total donation amounts and numbers of patrons were recorded. If the watching-eyes effect is replicable in this setting, then an image of eyes on the donation sign should yield higher donations than the non-social control condition (chair). Additionally, if it is indeed the monitoring function of the eyes that increases prosocial behavior, then eyes should elicit greater donations than other cues of human presence (mouth and nose).

Method

Setting and Participants

This study was conducted in a local children's museum in a small, mid-Atlantic university town in the United States. The museum caters to families with children up to 8 years of age. Overall, the demographics of visitors to the museum reflect those of the surrounding area. According to the US Census Bureau (2016), the median household income for the surrounding area is $49,775, the majority of residents are white (69.9% white; 19.0% black; 7.6% Asian; 3.5% other), and around half (49.8%) of the residents have obtained a college degree. During the 28 weeks in which data were collected, a total of 34,140 patrons (including adults and children) visited the museum and were exposed to the experimental manipulation of the donation box.

Procedure

The museum has a permanent, transparent donation box (without any accompanying images) with the sign "Donations would be appreciated" at the check-in desk at the front of the museum. Typically, patrons pay a price to enter the museum and in addition are welcome at any time to contribute to the donation box. All proceeds collected from the donation go toward the museum's general upkeep. For the purposes of this study, for a total of 28 weeks (between 2015 and 2017) the regular sign on the donation box was replaced by one of four signs that we provided the museum (pictures of the signs for all four conditions are provided in the ESM). These signs featured two 1 × 2-in. photographs of eyes, noses, mouths, or chairs. On any given sign, the two photographs showed different exemplars of the same item (e.g., one male and one female pair of eyes) above and below the phrase "Donations would be appreciated" (see Fig. 1 for an example from the Eyes condition). Pictures of the eyes, noses, and mouths were individually cropped from neutral-expression photos taken from the previously validated FACES database (Ebner et al. 2010). The pictures used in the chair condition were taken from the Bank of Standardized Stimuli (BOSS; Brodeur et al. 2014).

Once a week, the manager at the museum, who was blind to the hypotheses of the study, changed the donation sign according to a randomized order that we had provided, removed all of the donation money from the previous week to ensure that the donation box was empty at the start of each new week, and recorded the donation amount and number of patrons (adults and children). The donation box was located at its usual location at the check-in desk. Museum employees working at the check-in desk were unaware that the donation behavior and the experimental manipulation were being recorded. Data were not collected during weeks when donation behavior and attendance at the museum were subject to high flux, such as when a week included holidays on which the museum was closed (e.g., Thanksgiving and New Year's Day) or closures due to weather. Finally, over the course of the study, the museum underwent renovations for a period of 4 months. During this time, the museum was closed and no data were collected. To minimize any effects of the renovations on our data, we ensured that each condition was carried out a similar number of times before and after the renovations. A chi-square test confirmed that there were no significant differences in the frequency of each condition before and after the renovations ($p = .76$).

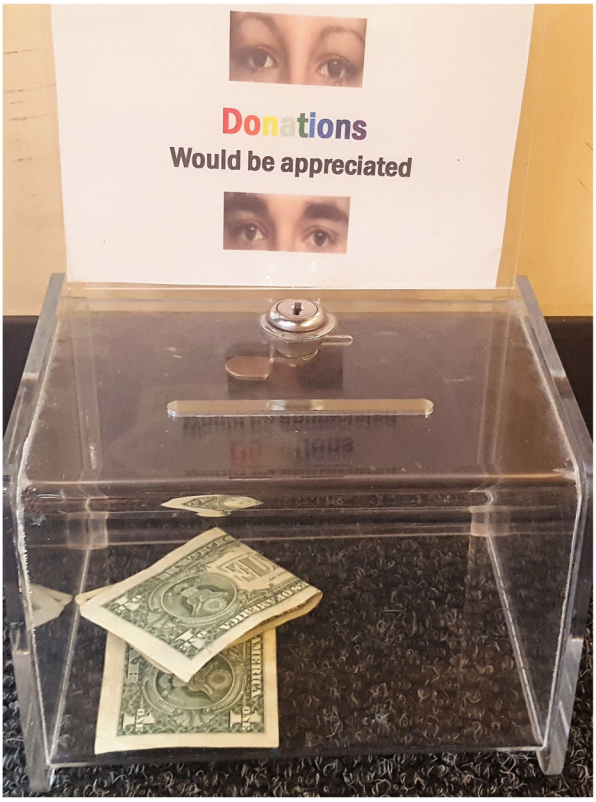


Fig. 1 Picture of the donation box with an example sign (from the Eyes condition)

Results

During the 28 weeks in which data were collected, the number of patrons who visited the museum per week (including adults and children) ranged from 533 to 1916 (M = 1219.29, SD = 406.24) and the total amount donated per week ranged from $3 to $46.33 ($M$ = $15.16, SD = $14.32). In order to account for the fluctuations in the number of patrons each week, we calculated the average amount of money donated per patron. An ANOVA with this average donation measure as the dependent variable and condition (Eyes, Nose, Mouth, Chair) as the independent variable revealed a significant effect of condition, $F_{3, 24} = 3.51$, $p = .031$, $\eta^2 = .31$ (Fig. 2).[1]

In relation to our first goal of replicating the watching-eyes effect, we compared the Eyes condition to the non-social (Chair) condition. A planned comparison using an independent samples t test indicated that the money donated per patron in the Eyes condition (M = .018, SD = .006) was significantly greater than in the Chair condition (M = .008, SD = .004), $t_{12} = 3.62$, $p = .004$, $d = 1.96$. We were thus able to replicate the basic watching-eyes effect.

[1] All of our results held when we used a one-way ANOVA with bootstrapping (10,000 replications) and performed post-hoc analyses with a Bonferroni correction.

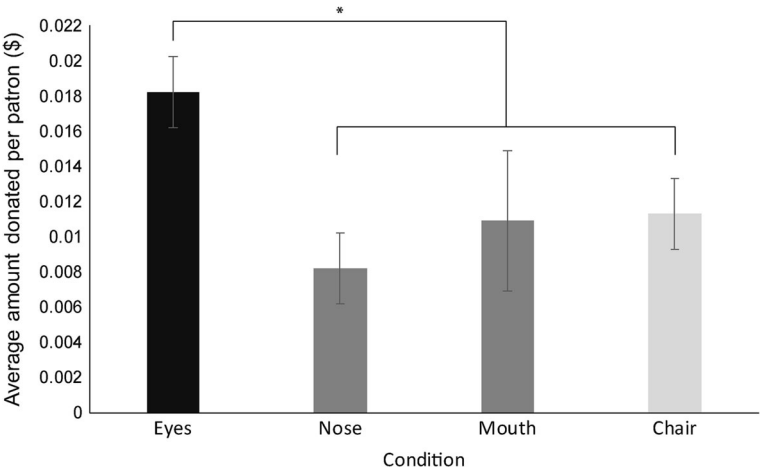


Fig. 2 The average amount of money donated per patron across conditions. Error bars indicate the standard error based on the total donation amounts for each week

In relation to our second goal of assessing whether the eyes increase donations more than other indicators of human presence, we pooled the two human-presence control conditions (Nose + Mouth, resulting in 14 weeks of data in total) and compared them with the Eyes condition. The donation per patron was greater in the Eyes condition ($M = .018$, $SD = .006$) than in the pooled human-presence control conditions ($M = .010$, $SD = .008$), $t_{19} = 2.55$, $p = .019$, $d = 1.13$.[2] When the Eyes condition was compared with each human-presence control condition separately, we found that the average amount donated per patron in the Eyes condition was greater than in the Nose condition ($M = .008$, $SD = .005$), $t_{12} = 3.29$, $p = .006$, $d = 1.81$. However, the average amount donated per patron in the Eyes condition was not significantly greater than in the Mouth condition ($M = .011$, $SD = .010$), $p = .13$.

Finally, exploratory analyses were conducted to examine the differences in donation behavior between the three control conditions (Nose, Mouth, Chair). There were no significant differences in donation behavior between the Nose condition and the Mouth condition, $p = .53$. In addition, donation behavior in the Chair condition did not significantly differ from either the Nose, $p = .90$, or the Mouth, $p = .47$, conditions.

Discussion

The current field study examined whether the presence of eyes enhances donation behavior more than a non-social cue and more than other social cues. By showing that participants exposed to eyes donated more than participants exposed to non-social visual cues (chair), the current study replicates previous research on the watching-eyes effect (Bateson et al. 2006; Haley and Fessler 2005). Our results further revealed greater donation behavior in response to the Eyes condition compared with other social

[2] This analysis had an unequal number of cells: 14 in Nose + Mouth pooled vs. 7 in the Eyes condition. We therefore also conducted a nonparametric test, which revealed a very similar result: Mann-Whitney U = 18.00, $p = .020$).

cues indexing human presence (Nose and Mouth conditions combined). The current findings therefore replicate and critically extend previous work and suggest that eyes play a special role in fostering cooperative behavior among humans. This is in line with the notion that cues to being watched result in reputation management behaviors (Bateson et al. 2006; Haley and Fessler 2005).

The watching-eyes effect has received a great deal of attention and has been demonstrated extensively (Bateson et al. 2006; Ernest-Jones et al. 2011; Haley and Fessler 2005). However, numerous recent studies, including meta-analyses, have failed to find strong support for the effect (Carbon and Hesslinger 2011; Fehr and Schneider 2010; Manesi and Pollet 2017; Northover et al. 2017; Saunders et al. 2016). We tested donation behavior in a real-world setting and were able to replicate the effect that watching eyes enhance prosocial behavior. Our results showed that the average donation amount per patron was greater in the presence of images of eyes than in a non-social control condition (images of chairs). Specifically, exposure to eyes yielded an average increase of $0.01 per patron compared with that of chairs. Although an increase of $0.01 per patron appears small, given the average number of museum patrons each week (more than 1200), this seemingly small difference amounts to an overall increase in donation of approximately $12 per week. Considering that the average donation amount per week (for the 28-week period) was about $15, a $12 increase is substantial and is more than what the results adjusted for patron number seem to suggest at first sight.

In light of conflicting reports in the literature, it is important to consider why the present study revealed an effect of watching eyes on prosocial behavior whereas some others have not. First, whereas some failed replications have employed stylized eyes (Fehr and Schneider 2010; Vogt et al. 2015), the eye stimuli in the present study were realistic photographs of human eyes, which might be more salient cues to being watched. Note, however, that prior studies using both stylized eyes (Haley and Fessler 2005; Vogt et al. 2015) and photographs of eyes (Bateson et al. 2006; Matsugasaki et al. 2015) have yielded mixed results. Thus, our use of photographs of eyes is unlikely to be the sole explanation for the discrepancy between our successful replication and prior non-replications.

Another possible explanation is based on findings indicating that context influences the amount donated in the presence of eye cues. In the present study, the donation box was located in the entrance area, which is typically busy with patrons checking in and leaving the museum, and a museum staff member was always present at the reception desk. These factors created a relatively public context for people's donations. The context is important because prior work indicates that attentiveness to eyes is predictive of donation behavior in a public (i.e., reputation-relevant) context but not when donation behavior is private or anonymous (Andreoni and Petrie 2004; Lamba and Mace 2010; Vaish et al. 2017). Accordingly, the public context in our study may have impacted and enhanced patrons' donation decisions, particularly in the presence of eyes. Moreover, since the donation box was transparent, patrons could see how much had already been donated, which might have resulted in participants inferring donation norms (see Fathi et al. 2014). This may have amplified the effect in the Eyes condition over the course of the week. Note, though, that this explanation has as its premise that more money was donated in the Eyes condition early in the week, which in turn provides support for the watching-eyes effect.

Finally, the positive result may be related to the fact that the current study was carried out in a real-world setting, in which, unlike in laboratory settings, participants do not know that they are part of an experiment. Indeed, positive results and larger effect sizes for watching eyes have been predominately obtained in field studies rather than in laboratory experiments. In fact, Northover et al. (2017), who did not find evidence for the watching-eyes effect, did not include some of the existing field studies in the first of two meta-analyses that they conducted—on the amount given after exposure to eyes (e.g., Ekström 2012; Powell et al. 2012). This indicates that there are contextual factors that may play a role in contributing to the watching-eyes effect. However, note that the second meta-analysis conducted by Northover et al. (2017)—on the proportion of participants who gave after exposure to eyes—did include some field studies (e.g., Ekström 2012), but still failed to find support for the watching-eyes effect. Clearly, more research is needed to determine whether and how contextual factors impact the watching-eyes effect.

The second goal of the present study was to better understand the mechanism underlying the watching-eyes effect. Prior work has generally assumed that the watching-eyes effect results from reputational concerns triggered by the monitoring function of the eyes (Bateson et al. 2006; Haley and Fessler 2005). However, an alternative possibility is that the eyes simply cue human presence, which increases prosociality for non-reputational reasons such as making individuals feel more social or part of a group, making them think about others' needs, or alerting them to possible sanctions for non-cooperation. In order to address this possibility, we compared the effect of eye cues on donation behavior with that of other human features (noses and mouths), which should similarly cue human presence. We reasoned that if reputational concern triggered by the monitoring function of the eyes is the key mechanism underlying the watching-eyes effect, then the eyes should be especially effective in enhancing donation behavior when compared with the other human features. In line with this, we found that patrons donated more money during weeks that featured the eyes than in the weeks that featured the other human features (Nose and Mouth pooled). These results are in agreement with prior (tentative) findings that watching eyes yield greater donations than closed eyes or eyes with an averted gaze (Manesi et al. 2016; see also Baillon et al. 2013). Together, these findings support the idea that the presence of eyes does not increase prosocial behavior simply by cueing social presence but rather through the monitoring function of the eyes, which enhances people's reputational concern and motivates them to engage in self-presentational behaviors.

One alternative to consider is that perhaps the eyes are more effective in cueing human presence than the other facial features. This is because when seen in isolation, they represent the most powerful trigger for face recognition processes (Keil 2009). In other words, eyes may be a sufficient feature to detect a face, whereas features such as noses and mouths may not by themselves be immediately recognized as faces and thus may not cue human presence as effectively. It will be informative for future work to examine how quickly and effectively the eyes versus other features are recognized as human stimuli and what impact this has on prosocial behavior.

We also explored potential differences in donation behavior between the Eyes condition and each individual human feature condition (Nose and Mouth) separately without specific a priori hypotheses for these comparisons. Donation behavior was significantly greater in the Eyes condition than the Nose condition but was not

significantly different between the Eyes and Mouth conditions. Importantly, donation behavior also did not significantly differ for the Mouth, Nose, and Chair conditions. The overall pattern with regard to the Mouth condition is difficult to interpret because no clear difference was obtained between the Mouth and any of the other conditions, leaving unclear whether the Mouth condition functioned more like the Eyes or more like the other control conditions (Nose and Chair). However, we might speculate that the mouth, which is critical for communicating with others and spreading information, might also be perceived as a (weaker) cue for reputation management and thus enhance donation behavior somewhat. These proposals are of course tentative and go beyond the conclusions that our data permit, but they raise interesting questions for future research.

In conclusion, the current study provides further empirical evidence that eyes enhance donation behavior in a real-life setting. Our results critically extend previous research by showing that viewing eyes selectively increases prosocial behavior, while viewing other cues that only indicate human presence does not have such an effect. These findings support accounts positing that eyes play a special role in fostering cooperative behaviors in humans, likely by eliciting reputation management.

Acknowledgments We are extremely grateful to the Virginia Discovery Museum in Charlottesville, VA, and in particular to Kaitlin Clear German for all her assistance with data collection. In addition, we would like to thank Janine Oostenbroek and Katie Krol for helpful discussions and comments on the manuscript.

References

Andreoni, J., & Petrie, R. (2004). Public goods experiments without confidentiality: A glimpse into fund-raising. *Journal of Public Economics, 88*(7), 1605–1623.

Baillon, A., Selim, A., & Van Dolder, D. (2013). On the social nature of eyes: The effect of social cues in interaction and individual choice tasks. *Evolution and Human Behavior, 34*(2), 146–154.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2*(3), 412–414.

Bateson, M., Callow, L., Holmes, J. R., Roche, M. L. R., & Nettle, D. (2013). Do images of "watching eyes" induce behaviour that is more pro-social or more normative? A field experiment on littering. *PLoS One, 8*(12), e82055.

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) phase II: 930 new normative photos. *PLoS One, 9*(9), e106953.

Bull, R., & Gibson-Robinson, E. (1981). The influences of eye-gaze, style of dress, and locality on the amounts of money donated to a charity. *Human Relations, 34*(10), 895–905.

Burnham, T. C., & Hare, B. (2007). Engineering human cooperation: Does involuntary neural activation increase public goods contributions? *Human Nature, 18*, 88–108.

Carbon, C. C., & Hesslinger, V. M. (2011). Bateson et al.'s (2006) cues-of-being-watched paradigm revisited. *Swiss Journal of Psychology, 70*, 203–210.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods, 42*(1), 351–362.

Ekström, M. (2012). Do watching eyes affect charitable giving? Evidence from a field experiment. *Experimental Economics, 15*(3), 530–546.

Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS One, 7*(10), e48433.

Ernest-Jones, M., Nettle, D., & Bateson, M. (2011). Effects of eye images on everyday cooperative behavior: A field experiment. *Evolution and Human Behavior, 32*(3), 172–178.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences, 99*(14), 9602–9605.

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences, 102*(47), 17245–17250.

Fathi, M., Bateson, M., & Nettle, D. (2014). Effects of watching eyes and norm cues on charitable giving in a surreptitious behavioral experiment. *Evolutionary Psychology, 12*(5), 878–887.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137–140.

Fehr, E., & Schneider, F. (2010). Eyes are on us, but nobody cares: Are eye cues relevant for strong reciprocity? *Proceedings of the Royal Society of London B: Biological Sciences, 277*(1686), 1315–1323.

Goffman, E. (1959). *The presentation of self in everyday life*. Garden City: Doubleday.

Grossmann, T. (2017). The eyes as windows into other minds: An integrative perspective. *Perspectives on Psychological Science, 12*, 107–121.

Haley, K. J., & Fessler, D. M. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior, 26*(3), 245–256.

Hare, B. (2017). Survival of the friendliest: *Homo sapiens* evolved via selection for prosociality. *Annual Review of Psychology, 68*, 155–186.

Keil, M. S. (2009). "I look in your eyes, honey": Internal face features induce spatial frequency preference for human face processing. *PLoS Computational Biology, 5*(3), e1000329.

Keller, J., & Pfattheicher, S. (2011). Vigilant self-regulation, cues of being watched and cooperativeness. *European Journal of Personality, 25*(5), 363–372.

Kurzban, R. (2001). The social psychophysics of cooperation: Nonverbal communication in a public goods game. *Journal of Nonverbal Behavior, 25*(4), 241–259.

Lamba, S., & Mace, R. (2010). People recognise when they are really anonymous in an economic game. *Evolution and Human Behavior, 31*(4), 271–278.

Leimgruber, K. L., Shaw, A., Santos, L. R., & Olson, K. R. (2012). Young children are more generous when others are aware of their actions. *PLoS One, 7*(10), e48292.

Manesi, Z., & Pollet, T. V. (2017). No support for the watching eyes effect across three "lost letter" field experiments. *Letters on Evolutionary Behavioral Science, 8*(1), 12–15.

Manesi, Z., Van Lange, P. A., & Pollet, T. V. (2016). Eyes wide open: Only eyes that pay attention promote prosocial behavior. *Evolutionary Psychology, 14*(2), 1–15.

Matsugasaki, K., Tsukamoto, W., & Ohtsubo, Y. (2015). Two failed replications of the watching eyes effect. *Letters on Evolutionary Behavioral Science, 6*(2), 17–20.

Milinski, M., Semmann, D., & Krambeck, H. (2002). Reputation helps solve the "tragedy of the commons." *Nature, 415*(6870), 424–426.

Nettle, D., Nott, K., & Bateson, M. (2012). 'Cycle thieves, we are watching you': Impact of a simple signage intervention against bicycle theft. *PLoS One, 7*(12), e51738.

Nettle, D., Harper, Z., Kidson, A., Stone, R., Penton-Voak, I. S., & Bateson, M. (2013). The watching eyes effect in the dictator game: it's not how much you give, it's being seen to give something. *Evolution and Human Behavior, 34*(1), 35–40.

Northover, S. B., Pedersen, W. C., Cohen, A. B., & Andrews, P. W. (2017). Artificial surveillance cues do not increase generosity: Two meta-analyses. *Evolution and Human Behavior, 38*(1), 144–153.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S., Breckler, S., et al. (2015). Promoting an open research culture. *Science, 348*(6242), 1422–1425.

Oda, R., Niwa, Y., Honma, A., & Hiraishi, K. (2011). An eye-like painting enhances the expectation of a good reputation. *Evolution and Human Behavior, 32*(3), 166–171.

Panagopoulos, C. (2014a). I've got my eyes on you: Implicit social pressure cues and prosocial behavior. *Political Psychology, 35*, 23–33.

Panagopoulos, C. (2014b). Watchful eyes: Implicit observability cues and voting. *Evolution and Human Behavior, 35*, 279–284.

Powell, K. L., Roberts, G., & Nettle, D. (2012). Eye images increase charitable donations: Evidence from an opportunistic field experiment in a supermarket. *Ethology, 118*(11), 1096–1101.

Saunders, T. J., Taylor, A. H., & Atkinson, Q. D. (2016). No evidence that a range of artificial monitoring cues influence online donations to charity in an MTurk sample. *Royal Society Open Science, 3*, 150710.

Sparks, A., & Barclay, P. (2013). Eye images increase generosity, but not for long: The limited effect of a false cue. *Evolution and Human Behavior, 34*(5), 317–322.

Sparks, A., & Barclay, P. (2015). No effect on condemnation of short or long exposure to eye images. *Letters on Evolutionary Behavioral Science, 6*(2), 13–16.

US Census Bureau. (2016). State and county QuickFacts: Charlottesville, Virginia. Retrieved from https://www.census.gov/quickfacts/fact/table/charlottesvillecityvirginiacounty/PST045216.

Vaish, A., Kelsey, C. M., Tripathi, A., & Grossmann, T. (2017). Attentiveness to eyes predicts generosity in a reputation-relevant context. *Evolution and Human Behavior, 38*(6), 729–733.

Vogt, S., Efferson, C., Berger, J., & Fehr, E. (2015). Eye spots do not increase altruism in children. *Evolution and Human Behavior, 36*(3), 224–231.

Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. *Science, 288*(5467), 850–852.

Caroline Kelsey obtained her BS in biology and psychology from Pennsylvania State University and her MA in psychology at The College of William and Mary. She is currently a doctoral candidate in developmental psychology at the University of Virginia. Her doctoral work has focused on the development of eye-based social cognition.

Amrisha Vaish earned her PhD in psychology from the Max Planck Institute for Evolutionary Anthropology and the Free University Berlin in Germany. She is an assistant professor in the Department of Psychology at the University of Virginia. Her research focuses on social and moral development, in particular the early emergence of prosocial behavior and its underlying motivations.

Tobias Grossmann obtained his PhD from the Max Planck Institute for Human Cognitive and Brain Sciences and the University of Leipzig in Germany. He is an associate professor in the Department of Psychology at the University of Virginia. His research is concerned with the developmental and brain origins of human social perception and cognition.

doi:10.1093/scan/nsr066 SCAN (2012) 7, 35–43

Facing puberty: associations between pubertal development and neural responses to affective facial displays

William E. Moore III,[1] Jennifer H. Pfeifer,[1] Carrie L. Masten,[2] John C. Mazziotta,[3] Marco Iacoboni,[3,4] and Mirella Dapretto[3,4]

[1]Department of Psychology, University of Oregon, [2]Center for Mind and Brain, University of California, Davis, [3]Ahmanson-Lovelace Brain Mapping Center, University of California, Los Angeles and [4]UCLA-FPR Center for Culture, Brain, and Development

Adolescence is marked by profound psychosocial and physiological changes. Although investigations into the interactions between these forces have begun to shed light on the neural correlates of affective processing during the transition to adolescence, relatively little is known about the relationship between pubertal development and emotion perception at the neural level. In the current longitudinal study, 45 neurotypical participants were shown affective facial displays while undergoing fMRI, at ages 10 and 13. Neural responses to emotional expressions at both time points were then correlated with a self-report measure of pubertal development, revealing positive associations with activity in amygdala, thalamus and visual cortical areas at age 10 that increased in magnitude and extent by age 13. At the latter time point, pubertal development was additionally correlated with enhanced responses to faces in temporal pole, ventrolateral prefrontal cortex (PFC) and dorsomedial PFC. Longitudinal comparisons revealed that the relationships between pubertal development and activity in the amygdala, hippocampus and temporal pole were significantly stronger during early adolescence than late childhood. These results suggest that pubertal development per se is linked to neural processing of socioemotional stimuli, particularly with respect to the integration of complex perceptual input and higher order cortical processing of affective content.

Keywords: adolescence; puberty; emotion; fMRI; amygdala; longitudinal

Folk wisdom suggests that teenagers are ruled by their emotions, in part due to their 'raging hormones'. Teens also have to cope with considerable interpersonal changes, as relationships with peers take on a new level of importance (Larson and Richards, 1991; Brown, 2004). Together, these facts raise questions concerning not only the adolescent emotional experience, but also how teens perceive emotional states in others (Dahl and Gunnar, 2009). In light of the 'social reorientation of adolescence' (Nelson *et al.*, 2005), reacting appropriately to the feelings of peers is an important skill for the developing brain to hone. Of course, in addition to shifting social and emotional contexts, adolescence is also a period of significant physiological change. While outward manifestations that signal the end of childhood are often plain to see (e.g. increased height), the effects of puberty on brain function are by nature more elusive. It is important to determine how neural responses to key socioemotional stimuli, such as affective facial displays, are affected by pubertal development—distinct from chronological age.

Although it has been said that the science of adolescent brain development is itself at an age often associated with pubescent growing pains (Steinberg, 2010), the field is in many ways still in its infancy. The neural networks associated with perceiving and reacting to emotional expressions have been targets of neuroimaging research for more than a decade (Adolphs, 2002; Haxby *et al.*, 2002), and patterns of structural and functional development in the regions comprising these networks (e.g. amygdala, prefrontal and extrastriate visual cortex) are garnering increasing scientific attention. Yet the relationships between pubertal and neural development have rarely been assessed directly in human adolescents (for exceptions see Silk *et al.*, 2009; Blakemore *et al.*, 2010; Forbes *et al.*, 2010, 2011).

Much of what *is* known about physiological changes in the brain during the transition from childhood to adulthood comes from research in animal populations, because adolescent behavior and neural function can be conserved across mammalian species (e.g. with respect to risk-taking; Sisk and Foster, 2004; Spear, 2004). Such homology is convenient to the extent that it allows for an enhanced degree of precision in generating hypotheses about pubertal contributions to

Received 27 April 2011; Accepted 16 September 2011

The authors wish to express their gratitude to Kristin McNealy, Larissa Borofsky, Nicole Vazquez, Elliot Berkman, Junaid Merchant and the University of Oregon Developmental Social Neuroscience and Social Affective Neuroscience Labs. For generous support the authors also wish to thank the Santa Fe Institute Consortium, Brain Mapping Medical Research Organization, Brain Mapping Support Foundation, Pierson-Lovelace Foundation, The Ahmanson Foundation, William M. and Linda R. Dietel Philanthropic Fund at the Northern Piedmont Community Foundation, Tamkin Foundation, Jennifer Jones-Simon Foundation, Capital Group Companies Charitable Foundation, Robson Family and Northstar Fund. The project described was supported by Grant Numbers RR12169, RR13642 and RR00865 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH); its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCR or NIH.

Correspondence should be addressed to Jennifer H. Pfeifer, Department of Psychology, 1227 University of Oregon, Eugene, OR 97403-1227, USA. E-mail: jpfeifer@uoregon.edu

© The Author (2012). Published by Oxford University Press. For Permissions, please email: journals.permissions@oup.com

Downloaded from http://scan.oxfordjournals.org/ at James Cook University on March 15, 2015

development. Approaches informed by empirical findings at the cellular and molecular levels are particularly relevant to the study of adolescence because they provide an appropriate analytic framework from which to consider the organizational and activational effects of hormones on the brain. For example, sex steroid hormones associated with typical development have been causally linked to changes in the hippocampus for social memory (Hebbard *et al.*, 2003) and reorganization of the visual cortex (Spear, 2009).

Although translational inference cannot speak directly to the role of hormones in human neural development, such findings can inform neuroimaging investigations in the selection and construction of homologous anatomical regions of interest (ROIs), for instance based on the distribution of endocrine receptors in the brains of other species (Goldstein *et al.*, 2001). Two such regions that have received considerable attention in developmental research are the prefrontal cortex (PFC) and the amygdala (Romeo, 2003; Sanz *et al.*, 2008; Schulz *et al.*, 2009). As it has been hypothesized that one of the principle consequences of pubertal increases in hormone levels is to reorganize motivational tendencies, specifically with respect to social targets (Forbes and Dahl, 2010), it seems likely that the PFC and amygala might be critical targets of the endocrine system during the transition to adolescence in humans as well. It is important to note, however, that the extension of conclusions derived from animal models to human adolescents must be undertaken with great care, as there are important differences between humans and other animals at levels of analysis ranging from the molecular (e.g. discrepancies in genetic makeup and expression) to the psychosocial (e.g. media influences on social behavior). Fortunately, advances in neuroimaging technology can help to bridge the gap between research in animal models and investigations of the living human brain.

Human neural development has been studied with both structural and functional magnetic resonance imaging (sMRI, fMRI) for years, but the literature is still relatively limited in scope. Here we briefly summarize known developmental trajectories that are relevant to the core network for processing affective facial displays, which includes both cortical (fusiform gyrus, superior temporal gyrus, post-central gyrus, temporal pole, anterior cingulate gyrus, insula, orbitofrontal cortex, PFC) and subcortical (brainstem, superior colliculus, lateral geniculate and pulvinar nuclei of the thalamus, hypothalamus, corpus striatum, amygdala, hippocampus) regions (Adolphs, 2002). Although total cerebral volume is nearly fully realized by the age of six, the proportions of different tissue types and relative size and composition of various anatomical structures undergo considerable change during the transition from childhood to adulthood (Lenroot and Giedd, 2010). White matter (WM) increases in a linear fashion throughout adolescence, and such increases imply enhanced connectivity between spatially remote areas of the cortex as well as between cortical and subcortical regions. The development of WM is most protracted in the PFC, a region which is crucial not only to response inhibition, planning and executive control (Behrens *et al.*, 2009), but also in computing the value of affective stimuli and determining appropriate behavior in social contexts (Sowell *et al.*, 2002; Rushworth *et al.*, 2007). Gray matter (GM), on the other hand, follows an inverted U-shaped developmental trajectory, with the inflection point occurring earliest in primary sensory cortex and latest in higher-order association cortices like lateral PFC and temporal pole (Giedd and Rapoport, 2010). Situated directly adjacent to the amygdala and orbitofrontal cortex, the temporal pole has been regarded as association cortex for affective perception, integrating sensory information about complex stimuli (e.g. facial expressions) with limbic signals indexing motivational salience and emotional content (Olson *et al.*, 2007). Along with PFC, GM reduction begins latest in the temporal poles, which are not fully mature until well into adulthood (Gogtay *et al.*, 2004).

This U-shaped developmental trajectory is common to a number of anatomical structures, but the exact point in time at which GM volume 'peaks' varies both across neural structures and across individuals (Lenroot and Giedd 2010). For example, the ventral striatum (vS) achieves maximal GM density during late childhood or early adolescence and subsequently undergoes a linear course of volumetric reduction (Sowell *et al.*, 2002). Central components of the limbic system like the amygdala and hippocampus exhibit slight increases in GM density into adulthood, but the rate of change plateaus in early adolescence, suggesting that the vast majority of their macroscopic development is complete shortly after a period of rapid expansion in late childhood (Ostby *et al.*, 2009).

In terms of neural function, there appears to be a similar U-shaped developmental trajectory of amygdalar response to emotional expressions, in which adolescents exhibit stronger amygdala reactivity upon perceiving emotional expressions than either children or adults (Guyer *et al.*, 2008). This has been explained in terms of a temporal mismatch: the subcortical structures are said to reach adult-like levels of functioning during early adolescence, while the relatively immature PFC has yet to achieve the ability to optimally regulate appetitive or aversive emotional signals (Casey *et al.*, 2010; Somerville *et al.*, 2010). For example, early reports posited that immature reward-value computation in the striatum was responsible for increases in adolescent risk-taking (Galvan *et al.*, 2006; Steinberg, 2008).

Recent longitudinal work has also demonstrated that striatal reactivity to emotional displays increases during the transition to adolescence, such that children exhibited increases in vS response from age 10 to 13 (as well as in vmPFC and temporal pole; Pfeifer *et al.*, 2011). Critically, this increase in vS activity was negatively correlated with susceptibility to peer influence and risky behavior, suggesting that the role of the vS in governing behavior is perhaps more complex than simply promoting impulsive approach motivation.

Downloaded from http://scan.oxfordjournals.org/ at James Cook University on March 15, 2015

There was also a longitudinal increase in negative coupling between vS and amygdala while viewing happy and sad facial expressions, relative to neutral ones. However, these findings did not address the influence of puberty on the development of adolescent neural functioning, in vS or elsewhere.

Therefore, an important next step is to simultaneously consider the coordinated activity of neural and endocrine systems in determining behavioral, emotional and psychosocial outcomes for teens (Dahl, 2011). The current investigation sought to address the influence of pubertal development on the neural correlates of affective processing during the transition to adolescence, rather than using age as a proxy for puberty. The ages of 10 and 13 are particularly relevant for addressing such questions, as they allow investigation of late childhood, when relatively few children have experienced significant pubertal development, and early adolescence, when most children have already progressed through the first several stages of puberty (Shirtcliff *et al.*, 2009). Furthermore, these ages represent important developmental epochs from a socioemotional perspective, including the end of elementary school and the end of middle school, respectively. Because data were collected during a tightly constrained window of chronological age during both late childhood and early adolescence, this sample provides a unique opportunity to study pubertal development in a continuous manner within a single timepoint, as well as to consider whether puberty exerts distinct effects in late childhood and early adolescence on the neural correlates of emotion reactivity. Our primary regions of interest, derived from the above literature review, included amygdala, PFC (both medial and lateral aspects), vS and temporal pole.

METHODS

Participants

Typically developing children ($N = 45$, 26 girls) completed data collection at two visits (separated by 36 ± 10 months), and provided high-quality fMRI and behavioral data at both timepoints (M's = 10.1 and 13.1 years, SDs = 0.30 and 0.31 years, at T1 and T2, respectively). Participants and their parents provided written informed assent/consent according to guidelines specified by the Institutional Review Board at UCLA. Participants had no history of significant medical, psychiatric, or neurological disorders.

The Pubertal Development Scale (PDS; Petersen *et al.*, 1988) was completed by participants at both timepoints. On this measure, participants self-report visible development of secondary sexual characteristics. The PDS was always administered with a researcher nearby to answer any questions participants had about the meaning of items. Parents were notified several days in advance of the session that a form assessing pubertal development would be part of the assessment, which gave parents the opportunity to talk with their children about this topic in advance if they so desired. We utilized the scoring guidelines provided by Crockett and colleagues (Petersen *et al.*, 1988), and we omitted the social comparison question (about development relative to the pace of peers). Across all participants, Cronbach's test indicated that the PDS was reliable (five items, $\alpha = 0.80$). There was a highly significant increase from T1 to T2 on the PDS (M's = 1.60 and 2.60, SDs = 0.37 and 0.62 at T1 and T2, respectively; $t(1,44) = 11.36$, $P \sim 0$). At T1, there was no relationship between chronological age and pubertal development [$r(43) = 0.11$, $P = 0.47$]. However, at T2, PDS scores and chronological age were moderately correlated [$r(43) = 0.34$, $P < 0.05$]. See Figure 1 for visualization of raw PDS data at both timepoints. Because there was no significant difference in mean PDS scores for male and female subjects at T1 [$t(1,43) = 0.25$, $P = 0.8$] or T2 [$t(1,43) = 1.55$, $P = 0.13$], fMRI analyses collapsed across gender to increase statistical power.

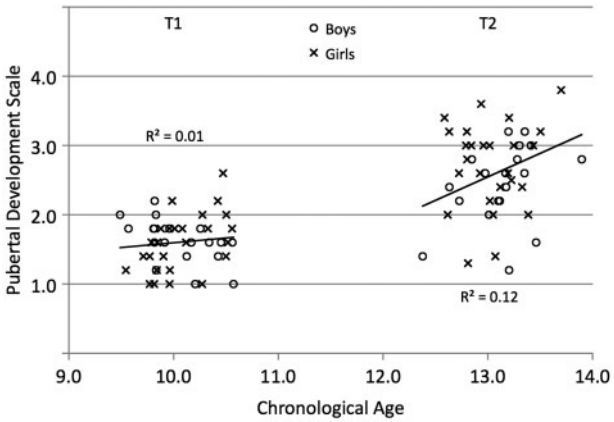


Fig. 1 Correlations between pubertal development and chronological age. Figure 1 depicts the relationship between pubertal development and chronological age at T1 [$r(43) = 0.11$, $P = 0.47$] and at T2 [$r(43) = 0.34$, $P < 0.05$].

Paradigm

During the fMRI scan, participants passively observed full-color, whole-face emotional displays (angry, fearful, happy, sad and neutral) from the NimStim set (Tottenham *et al.*, 2009). Events lasted 2 s, with an interstimulus interval of variable (jittered) length ranging from 0.5 to 1.5 s ($M = 1$ s), and were presented in counterbalanced orders optimized for efficient detection of contrasts between emotions using a genetic algorithm (Wager and Nichols, 2003). A total of 96 whole-brain volumes were acquired on a Siemens Allegra 3.0 Tesla MRI scanner at each time point, including the 80 stimuli described above and an additional 16 null events (fixation crosses at eye-level).

fMRI acquisition and analysis

Data were acquired using a Siemens Allegra 3.0 Tesla MRI scanner. A 2D spin-echo scout (TR = 4000 ms, TE = 40 ms, matrix size 256 by 256, 4-mm thick, 1-mm gap) was acquired in the sagittal plane to allow prescription of the slices to be obtained in the remaining scans.

Downloaded from http://scan.oxfordjournals.org/ at James Cook University on March 15, 2015

The scan lasted 4 min and 54 s (gradient-echo, TR = 3000 ms, TE = 25 ms, flip angle = 90°, matrix size 64 by 64, FOV = 20 cm, 36 slices, 3.125-mm in-plane resolution, 3-mm thick). For each participant, a high-resolution structural T2-weighted echo-planar imaging volume (spin-echo, TR = 5000 ms, TE = 33 ms, matrix size 128 by 128, FOV = 20 cm, 36 slices, 1.56-mm in-plane resolution, 3-mm thick) was also acquired coplanar with the functional scan. Stimuli were presented to participants through high-resolution magnet-compatible goggles (Resonance Technology, Inc.).

Using Automated Image Registration (Woods *et al.*, 1999) implemented in the LONI Pipeline Processing Environment (http://www.pipeline.loni.ucla.edu; Rex *et al.*, 2003), all functional images were (i) realigned to correct for head motion and co-registered to their respective high-resolution structural images using a six-parameter rigid body transformation model, (ii) spatially normalized into a Talairach-compatible MR atlas using polynomial non-linear warping and (iii) smoothed using a 6-mm FWHM isotropic Gaussian kernel. No participant averaged >2.0 mm of motion at either time point (*M*'s = 0.317 and 0.413 mm, SDs = 0.262 and 0.365, at T1 and T2, respectively), and no participant moved >2.0 mm between any image. All coordinates are reported in Talairach space.

Statistical analyses were implemented in SPM8 (Wellcome Department of Cognitive Neurology, London, UK; http://www.fil.ion.ucl.ac.uk/spm/). For each subject, condition effects were estimated according to the general linear model, using a canonical hemodynamic response function, high-pass filtering (128 s), AR(1) and no global scaling. At the fixed effects (single subject) level, linear contrasts were constructed to assess comparisons of interest within individual participants at each timepoint (each of the expressions *vs* null events as well as the average of all expressions *vs* null events). In other words, activity for each emotional expression was contrasted against a resting baseline rather than a low-level control. Because neutral facial expressions are socially meaningful and often elicit considerable activity in regions associated with processing affective salience and uncertainty, results comparing emotional expressions to neutral ones may be difficult to interpret (Kober *et al.*, 2008; Pfeifer *et al.*, 2011). An ideal control is one that possesses the visual complexity of facial expressions with no social or affective content (e.g. scrambled faces; Rahko *et al.*, 2010). Unfortunately, scrambled faces were not included in this paradigm at the outset of the longitudinal investigation. However, comparisons between each facial expression and null events allowed us to identify where neural responses to affective displays were more or less associated with pubertal development (a continuous predictor) across subjects, independent of mean levels of activity. Furthermore, in the group level analyses we could then contrast any of the five facial expressions versus rest, as well as the four emotional expressions versus neutral.

At the group level, we used two complementary analysis strategies. First, in order to specifically examine the effects of puberty on reactivity to all affective displays as well as specific facial expressions at T1 and T2, separate random effects analyses were based on the contrast images from each expression, using each participant's PDS score as the sole predictor. Second, a repeated measures mixed design was employed in order to quantify longitudinal change in neural activity to facial expressions that was associated with pubertal development. In order to assess the influence of puberty independently from that of chronological age, linear regression was conducted to obtain unstandardized residual values of pubertal development, indexing variability in PDS scores that could not be explained by age (in other words, age was used as a covariate in the regression computed to obtain the residualized PDS scores). Including the residualized PDS scores as regressors in this design, with time and emotion as factors, replicated previous reports of longitudinal increases in vS activity for all expressions compared to rest (see Pfeifer *et al.*, 2011, Table 1), despite slight changes in the sample to achieve greater gender balance (increasing proportion of boys and decreasing proportion of girls). Correlations between the BOLD signal elicited by the task and behavioral measures of social, affective and interpersonal functioning used in prior studies conducted on overlapping subsets of participants (Interpersonal Reactivity Index, Davis, 1983; Interpersonal Competence Scale, Cairns *et al.*, 1996; Resistance to Peer Influence, Steinberg and Monahan, 2007; and Indicators of Risk Behavior and Delinquency, Gestsdóttir and Lerner, 2007) did not reveal any significant relationships beyond those reported previously (Pfeifer *et al.*, 2008, 2011). For all whole-brain analyses, results were reported at an alpha threshold of $p < 0.05$ for magnitude at peak voxels, corrected for multiple comparisons using False Discovery Rate (Genovese *et al.*, 2002), and an additional extent threshold of $p < 0.05$ at the cluster level, derived using a Monte Carlo simulation (3dClustSim, the successor to AlphaSim in the AFNI software package; Cox, 1996).

RESULTS

Based on the suite of regression analyses, positive correlations with pubertal development at age 10 were constrained to extrastriate cortex, thalamus and amygdala, while by age 13, level of pubertal development was further correlated with increased activity in the left temporal pole, ventrolateral PFC and ventromedial PFC for the combination of all affective facial displays (see Figure 2; for comprehensive reports of additional activity correlated with pubertal development at each timepoint, see Tables 1 and 2).

Similar patterns were observed when each expression was queried individually, although there were some notable differences. At age 10, pubertal development was correlated with greater activity in extrastriate cortex and thalamus in response to angry, fearful, happy and neutral facial

Downloaded from http://scan.oxfordjournals.org/ at James Cook University on March 15, 2015

Table 1 Positive correlations between pubertal development and neural response to affective facial displays at Time 1

Time 1	Region		*k*	*t*	*r*	*x*	*y*	*z*
All expressions	Extrastriate cortex	BA 18	7011	10.26	0.71	2	−84	2
	Extrastriate cortex	BA 18		9.26	0.66	−14	−94	12
	Thalamus		116	6.29	0.47	18	−32	4
	Thalamus			5.50	0.41	−22	−32	2
	Amygdala		108	4.59	0.32	−22	−12	−12
Anger	Extrastriate cortex	BA 18	4160	6.80	0.51	−20	−80	−8
Fear	Extrastriate cortex	BA 18	6513	11.12	0.74	0	−80	4
	Thalamus		160	5.98	0.45	18	−30	4
	Thalamus		117	5.24	0.38	−20	−32	2
Happy	Extrastriate cortex	BA 18	6302	10.98	0.86	6	−90	12
	Extrastriate cortex	BA 18		10.94	0.86	−18	−76	−6
	Thalamus		99	5.61	0.65	16	−32	4
	Thalamus		146	5.57	0.64	−24	−32	2
Sad	Striate cortex	BA 17	6052	12.81	0.89	8	−88	2
Neutral	Extrastriate cortex	BA 18	6057	11.07	0.86	−14	−94	12
	Extrastriate cortex	BA 18		9.89	0.83	4	−92	16
	Thalamus		107	6.06	0.67	18	−30	4
	Amygdala		113	3.61	0.48	26	−4	−12

Table 1 describes neural responses that are significantly correlated with pubertal development at T1. *k* corresponds to the spatial extent of significant clusters (2 mm^3 voxels). *t* represents the *t*-statistic at the peak voxel. *r* indexes the correlation between PDS score and activity at the peak voxel. *x*, *y* and *z* refer to the Talairach coordinates corresponding to the left-right, anterior-posterior and inferior-superior axes, respectively. All activity is reported at a peak threshold of $P < 0.05$ (FDR corrected) and a cluster threshold of $P < 0.05$.

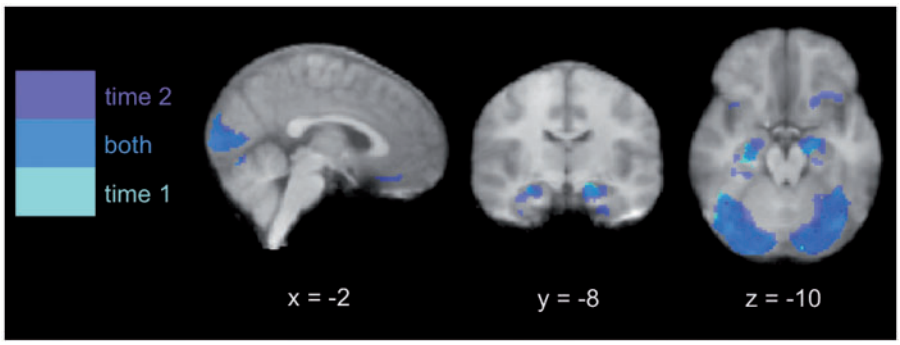


Fig. 2 Correlations between neural activity and pubertal development. Figure 2 depicts activity in extrastriate cortex, amygdala, and hippocampus that correlates with pubertal development at T1 and T2. x, y and z refer to the Talairach coordinates corresponding to the left-right, anterior-posterior and inferior-superior axes, respectively. All activity is displayed at a peak threshold of $P < .05$ (FDR corrected) and a cluster threshold of $P < 0.05$.

expressions. This pattern was even more robust at age 13, and also included the neural response to sad expressions. At 10 years of age, activation in the amygdala was only modulated by pubertal development while viewing neutral facial expressions. However, at age 13, level of pubertal development was associated with significantly greater activation in the amygdala for all five expressions examined, although whether the response was lateralized to the left or right hemispheres, or bilateral, varied by emotion. Furthermore, at age 10 there were no correlations in the temporal pole between pubertal development and neural response to affective facial displays, but by age 13 pubertal development was significantly linked to enhanced activity in temporal pole for happy, sad, fearful and neutral expressions. Positive correlations between pubertal development and ventrolateral PFC response for fearful, neutral and happy expressions were also evident at age 13 but not age 10. Finally, this pattern was replicated in dorsomedial PFC responses, but only for happy facial expressions.

Although there were no statistically significant differences in PDS scores between male and female participants, confirmatory fMRI analyses were carried out to ensure that gender was not a confound. Direct statistical comparisons including gender as a factor revealed no significant differences in the relationship between pubertal development and neural response to facial expressions. However, conducting separate linear regression analyses by gender revealed slight differences in cluster sizes, stereotactic locations and *t*-values. For example, for the average of all expressions compared to rest at T2, activity in ventrolateral PFC was bilateral for girls,

Downloaded from http://scan.oxfordjournals.org/ at James Cook University on March 15, 2015

Table 2 Positive correlations between pubertal development and neural response to affective facial displays at Time 2

Time 2	Region		*k*	*t*	*r*	*x*	*y*	*z*
All expressions	Extrastriate cortex	BA 18	9713	10.62	0.72	8	−82	−2
	Fusiform gyrus	BA 37		9.52	0.67	36	−48	−14
	Thalamus			6.72	0.51	−24	−30	2
	Thalamus			6.38	0.48	18	−30	4
	Amygdala		691	4.82	0.35	−24	−12	−12
	Amygdala	BA 28	667	4.70	0.33	14	−4	−12
	Temporal pole	BA 38		4.37	0.30	−38	14	−22
	vlPFC	BA 47	119	3.93	0.26	−40	20	−12
	vlPFC	BA 47	94	3.84	0.25	28	28	−12
	vmPFC	BA 11		4.25	0.29	20	24	−10
Anger	Extrastriate cortex	BA 18	7821	10.53	0.72	8	−82	0
	Cuneus	BA 17		9.89	0.69	−14	−90	4
	Thalamus		121	6.24	0.47	−24	−32	2
	Thalamus		137	5.64	0.42	18	−30	4
	Amygdala		117	5.41	0.40	16	−6	−10
Fear	Extrastriate cortex	BA 18	8830	10.65	0.72	8	−80	0
	Extrastriate cortex	BA 18	241	10.48	0.71	−10	−80	0
	Thalamus		313	5.38	0.40	20	−28	2
	Amygdala		69	3.72	0.24	24	−4	−14
	Amygdala		416	4.40	0.31	−24	−2	−12
	Fusiform gyrus	BA 20		4.39	0.30	−38	−12	−22
	Temporal pole	BA 38		4.55	0.32	−36	14	−22
	vlPFC	BA 47	53	3.94	0.26	−40	22	−12
	vlPFC	BA 47	47	3.68	0.24	40	28	−8
Happy	Extrastriate cortex	BA 18	9062	10.39	0.84	8	−78	−2
	Thalamus		244	5.49	0.64	18	−30	2
	Thalamus		134	4.41	0.55	−22	−30	0
	Temporal pole	BA 38	140	5.45	0.63	−40	16	−24
	Temporal pole	BA 38	402	4.69	0.58	32	10	−30
	Amygdala			4.30	0.54	26	−6	−14
	Amygdala		203	4.77	0.58	−20	−8	−14
	dmPFC	BA 10	104	3.74	0.49	4	56	18
	vlPFC	BA 47		3.85	0.50	−40	20	−12
Sad	Extrastriate cortex	BA 18	9748	12.41	0.88	6	−80	0
	Thalamus		262	5.85	0.66	−24	−30	0
	Thalamus		802	6.98	0.72	24	−32	4
	Temporal pole	BA 38		5.56	0.64	30	12	−30
	Temporal pole	BA 38	76	4.31	0.54	−40	16	−20
	Amygdala		198	4.01	0.52	−24	−10	−10
	Amygdala			3.67	0.48	−32	−8	−14
Neutral	Extrastriate cortex	BA 18	9252	11.70	0.87	6	−82	0
	Striate cortex	BA 17		10.32	0.84	−20	−90	4
	Thalamus		260	5.33	0.63	−26	−30	0
	Thalamus		118	4.45	0.56	20	−30	2
	Temporal pole	BA 38	119	4.83	0.59	−38	14	−24
	vlPFC	BA 47		3.99	0.52	−42	20	−14
	Hippocampus		132	4.11	0.53	−24	−14	−12
	Amygdala			4.07	0.52	−28	−6	−16
	Amygdala		94	3.26	0.44	26	−8	−12

Table 2 describes neural responses that are significantly correlated with pubertal development at T2. *k* corresponds to the spatial extent of significant clusters (2 mm^3 voxels). *t* represents the *t*-statistic at the peak voxel. *r* indexes the correlation between PDS score and activity at the peak voxel. *x*, *y* and *z* refer to the Talairach coordinates corresponding to the left-right, anterior-posterior, and inferior-superior axes, respectively. All activity is reported at a peak threshold of $P < 0.05$ (FDR corrected) and a cluster threshold of $P < 0.05$.

but left lateralized for boys. Such minor differences may have resulted from slight variability between groups in sample size.

Looking to the repeated measures mixed design, no significant relationship between pubertal development (independent of age) and BOLD response to emotional expressions versus neutral ones was apparent at T1. However, at T2 pubertal development (independent of age) was correlated with activity in visual cortex, right amygdala, bilateral hippocampus and temporal pole. These patterns, identified by querying each timepoint individually,

Downloaded from http://scan.oxfordjournals.org/ at James Cook University on March 15, 2015

were statistically different from each other: the positive correlations between pubertal development (independent of age) in visual cortex ([−14, −94, 12], $t = 9.81$), bilateral hippocampus (left: [−24, −32, 4], $t = 4.22$, right: [20, −30, 4], $t = 5.37$), bilateral amygdala (left: [−24, −12, −12], $t = 3.22$, right: [18, −6, −10], $t = 4.47$) and bilateral temporal pole (left: [−42, 16, −12], $t = 3.55$, right: [32, −8, 26], $t = 3.61$) were significantly greater at T2 than T1, as demonstrated by direct comparison of the two timepoints within a repeated measures mixed design. These results imply an interaction between time and pubertal development, insofar as the effects of pubertal development are significantly different between the two waves of data collection. It should be reiterated that this statistical approach used residualized PDS values in order to examine the unique effect of pubertal development, independent from that of age.

DISCUSSION

These results provide evidence of changing relations between pubertal development and neural responses to affective facial displays across the transition from late childhood to early adolescence. Recent longitudinal work focusing on this specific period (Pfeifer *et al.*, 2011) demonstrated age-related increases in a number of key regions of interest, particularly the vS. Here, a different subset of regions exhibited longitudinal increases with respect to pubertal development, including the amygdala and lateral PFC. This is consistent with the idea that puberty per se, and not just age, is associated with enhanced reactivity to socioemotional stimuli, as would be expected in the 'social reorientation of adolescence' (Nelson *et al.*, 2005). Here, neural activity in the amygdala, hippocampus and temporal pole elicited by the affective content of facial displays was significantly correlated with an age-independent measure of pubertal development by age 13, accompanied by significant longitudinal increases in the strength of this coupling.

Our prior longitudinal fMRI study (Pfeifer *et al.*, 2011) observed that, unlike vS reactivity to emotional expressions, the amygdala does not exhibit a significantly greater response at age 13 than at age 10 (except to sad faces). However, we now show that pubertal development enhances amygdala responses to emotions in late childhood, and this link appears to strengthen in early adolescence such that the two are correlated across all expressions individually at this time (happy, neutral, sad, fearful and angry). Activity in the thalamus and extrastriate cortex, structures that work in tandem with the amygdala to process the affective salience of social signals (Adolphs, 2008), shows a similar pattern (although visual cortical regions also exhibit increase over time in the absence of pubertal predictors). These data begin to elucidate the neural foundations of how puberty may 'reorient' adolescents toward social stimuli, such that complex communicative signals like emotional facial expressions begin to take increased relevance as children transition into their teenage years.

Recent findings on social threat and pubertal development have indicated that activity in amygdala and ventrolateral PFC while viewing fearful or neutral faces is stronger during early stages of pubertal development than later stages (Forbes *et al.*, 2011), suggesting that a reduction in threat-related response may be adaptive for adolescents insofar as promoting social approach. The opposite patterns elucidated in the present study may be attributable to differences in fMRI data analysis (e.g. whole brain regression as opposed to ROI analysis), or differences in manner of calculating pubertal development using the PDS. These differences may also be due to the fact that our design employed both longitudinal and cross sectional approaches, or the fact that we also used non-threatening facial expressions such as happiness and sadness.

We discerned another interesting pattern in the responses of several regions that are not significantly related to puberty at age 10, but tightly linked to it at age 13, including the medial PFC, lateral PFC and temporal pole. By early adolescence, the neural response to emotional expressions in these regions is positively correlated with pubertal development. This suggests puberty may affect frontal and temporal association cortices in a way that allows for increasingly mature integration of perceptual and affective information. These higher order cortical areas will continue to develop well into adulthood, and while their functional profile differs for adolescents (Casey *et al.*, 2005; Burnett and Blakemore, 2009; Pfeifer *et al.*, 2009), it is difficult to draw clear boundaries for what constitutes 'mature activity' in any isolated region (Poldrack, 2010). Because adult brain activity is the product of a large number of regions and systems working in concert, the 'maturity' of a functional neural response cannot be fully assessed independently from that of other regions. Thus it may be helpful to frame adolescent neural development in terms of network-wide changes including not only higher-order association cortices, but also their interactions with subcortical regions of interest.

Our findings also demonstrate that different anatomical structures have different developmental profiles. For example, the relationship between puberty and change in amygdala activity appears relatively continuous, as it correlates with pubertal development at both timepoints. The transition from late childhood to early adolescence, however, cannot be explained purely in linear terms. Several important discrete events are also likely to take place within this age range, such as entry into middle school, the start of early romantic relationships, and a shift in self-concept from regarding oneself as a child to internally adopting the mantle of teenager (Pfeifer *et al.*, 2009). However, the relationship between pubertal development and activity in ventrolateral PFC, medial PFC and temporal pole is likely driven by both developmental stage and puberty. The relationship between puberty and responses to emotional stimuli in temporal pole (longitudinal increases which also correlated with increases in resistance to peer influence;

Downloaded from http://scan.oxfordjournals.org/ at James Cook University on March 15, 2015

Pfeifer *et al.*, 2011) in particular speaks to a pattern by which the successful integration of perceptual input from complex social stimuli is gradually incorporated with online emotional processing and inhibitory control. This process is more complex than a simple dichotomy between the roles of subcortical and cortical structures is suited to explain, especially in light of the neural consequences of puberty.

While this work is an important step toward studying adolescence in a more comprehensive manner, it is only a preliminary one. Relationships established in the current work between pubertal development and changes in neural activity of key regions like the amygdala and temporal pole may help lay the foundation for more comprehensive study of the way that the adolescent endocrine system influences the roles of these and other structures in perceiving and experiencing emotion. An ideal advancement would be to simultaneously investigate pubertal changes at multiple levels of analysis (Cacioppo and Berntson, 1992), such as by measuring hormones and administering physical exams in addition to self-report measures and neuroimaging techniques. Also much needed in the field of adolescent neuroscience is the application of methods that allow for data driven, model testing approaches, which will allow investigators to more robustly account for the large number of variables at work. As our understanding of the biological mechanisms underlying adolescent emotion processing grows, we will be better equipped to address the mental health problems, risk-taking behavior and profound social changes that often accompany this developmental epoch. By mapping out the dynamic systems involved in the transition from childhood through adolescence to adulthood, we will be able to better understand how social context, pubertal hormones and neural function simultaneously affect adolescent motivational tendencies and emotional experiences, bringing us closer to unlocking the mysteries of the teenage brain.

REFERENCES

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral Cognitive Neuroscience Review*, *1*(1), 21–62.

Adolphs, R. (2008). Fear, faces, and the human amygdala. *Current Opinion in Neurobiology*, *18*(2), 166–72.

Behrens, T.E., Hunt, L.T., Rushworth, M.F. (2009). The Computation of Social Behavior. *Science* 324(5931) (2009): doi:10.1126/science.1169694.

Brown, B.B. (2004). Adolescents' relationships with peers. *Handbook of Adolescent Psychology*, *2*, 363–94.

Burnett, S., Blakemore, S.J. (2009). Functional connectivity during a social emotion task in adolescents and in adults. *The European Journal of Neuroscience*, *29*(6), 1294–301.

Cacioppo, J.T., Berntson, G.G. (1992). Social psychological contributions to the decade of the brain. Doctrine of multilevel analysis. *The American Psychologist*, *47*(8), 1019–28.

Cairns, R.B., Leung, M.C., Gest, S.D., Cairns, B.D. (1995). A brief method for assessing social development: structure, reliability, stability, and developmental validity of the Interpersonal Competence Scale. *Behavioral Research and Therapy*, *33*, 725–36.

Casey, B.J., Duhoux, S., Cohen, M.M. (2010). Adolescence: what do transmission, transition, and translation have to do with it? *Neuron*, *67*(5), 749–60.

Casey, B.J., Galvan, A., Hare, T.A. (2005). Changes in cerebral functional organization during cognitive development. *Current Opinion in Neurobiology*, *15*(2), 239–44.

Casey, B.J., Jones, R.M., Somerville, L.H. (2011). Braking and accelerating of the adolescent brain. *Journal of Research on Adolescence*, *21*(1), 21–33.

Cohen, J.R., Asarnow, R.F., Sabb, F.W., et al. (2010). A unique adolescent response to reward prediction errors. *Nature Neuroscience*, *13*(6), 669–71.

Cox, R.W. (1996). AFNI: software analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–73.

Dahl, R. (2011). Understanding the risky business of adolescence. *Neuron*, *69*(5), 837–9.

Dahl, R.E., Gunnar, M.R. (2009). Heightened stress responsiveness and emotional reactivity during pubertal maturation: implications for psychopathology. *Development and Psychopathology*, *21*(01), 1–6.

Davis, M.H. (1983). Measuring individual differences in empathy: evidence for a multi-dimensional approach. *Journal of Personality and Social Psychology*, (44), 113–26.

Delgado, M.R., Li, J., Schiller, D., Phelps, E.A. (2008). The role of the striatum in aversive learning and aversive prediction errors. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1511), 3787–800.

Forbes, E.E., Dahl, R.E. (2010). Pubertal development and behavior: hormonal activation of social and motivational tendencies. *Brain and Cognition*, *72*(1), 66–72.

Forbes, E.E., Ryan, N.D., Phillips, M.L., et al. (2010). Healthy adolescents' neural response to reward: Associations with puberty, positive affect, and depressive symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(2), 162–72.

Forbes, E.E., Phillips, M.L., Ryan, N.D., Dahl, R.E. (2011). Neural systems of threat processing in adolescents: role of pubertal maturation and relation to measures of negative affect. *Developmental Neuropsychology*, *36*(4), 429–52.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301.

Galvan, A., Hare, T.A., Parra, C.E., et al. (2006). Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(25), 6885–92.

Genovese, C.R., Lazar, N.A., Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, *15*(4), 870–8.

Gestsdóttir, S., Lerner, R.M. (2007). Intentional self-regulation and positive youth development in early adolescence: findings from the 4-H Study of Positive Youth Development. *Developmental Psychology*, *43*, 508–21.

Giedd, J.N., Rapoport, J.L. (2010). Structural MRI of pediatric brain development: What have we learned and where are we going? *Neuron*, *67*(5), 728–34.

Gogtay, N., Giedd, J.N., Lusk, L., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(21), 8174.

Goldstein, J.M., Seidman, L.J., Horton, N.J., et al. (2001). Normal sexual dimorphism of the adult human brain assessed by in vivo magnetic resonance imaging. *Cerebral Cortex (New York, N.Y. : 1991)*, *11*(6), 490.

Guyer, A.E., Monk, C.S., McClure-Tone, E.B., et al. (2008). A developmental examination of amygdala response to facial expressions. *Journal of Cognitive Neuroscience*, *20*(9), 1565–82.

Hare, T.A., Tottenham, N., Galvan, A., Voss, H.U., Glover, G.H., Casey, B.J. (2008). Biological substrates of emotional reactivity and regulation in adolescence during an emotional go-nogo task. *Biological Psychiatry*, *63*(10), 927–34.

Downloaded from http://scan.oxfordjournals.org/ at James Cook University on March 15, 2015

Haxby, J.V., Hoffman, E.A., Gobbini, M.I. Human neural systems for face recognition and social communication. *Biological psychiatry*, *51*(1), 59–67.

Hebbard, P.C., King, R.R., Malsbury, C.W., Harley, C.W. (2003). Two organizational effects of pubertal testosterone in male rats: Transient social memory and a shift away from long-term potentiation following a tetanus in hippocampal CA1. *Experimental Neurology*, *182*(2), 470–475.

Kober, H., Barrett, L.F., Joseph, J., Bliss-Moreau, E., Lindquist, K., Wager, T.D. (2008). Functional grouping and cortical–subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *Neuroimage*, *42*(2), 998.

Larson, R., Richards, M.H. (1991). Daily companionship in late childhood and early adolescence: changing developmental contexts. *Child Development*, *62*(2), 284–300.

Lenroot, R.K., Giedd, J.N. (2010). Sex differences in the adolescent brain. *Brain and Cognition*, *72*(1), 46–55.

McClure, S.M., York, M.K., Montague, P.R. (2004). The neural substrates of reward processing in humans: the modern role of fmri. *The Neuroscientist*, *10*(3), 260.

Olson, I.R., Plotzker, A., Ezzyat, Y. (2007). The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain : A Journal of Neurology*, *130*(Pt 7), 1718–31.

Ostby, Y., Tamnes, C.K., Fjell, A.M., Westlye, L.T., Due-Tønnessen, P., Walhovd, K.B. (2009). Heterogeneity in subcortical brain development: a structural magnetic resonance imaging study of brain maturation from 8 to 30 years. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(38), 11772–82.

Petersen, A.C., Crockett, L., Richards, M., Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence*, *17*(2), 117–33.

Pfeifer, J.H., Iacoboni, M., Mazziotta, J.C., Dapretto, M. (2008). Mirroring others' emotions relates to empathy and social abilities during childhood. *NeuroImage*, *39*, 2076–85.

Pfeifer, J.H., Masten, C.L., Borofsky, L.A., Dapretto, M., Fuligni, A.J., Lieberman, M.D. (2009). Neural correlates of direct and reflected self-appraisals in adolescents and adults: When social perspective-taking informs self-perception. *Child Development*, *80*(4), 1016–38.

Pfeifer, J.H., Masten, C.L., Moore, W.E., et al. (2011). Entering adolescence: resistance to peer influence, risky behavior, and neural changes in emotion reactivity. *Neuron*, *69*(5), 1029–36.

Phillips, M.L., Drevets, W.C., Rauch, S.L., Lane, R. (2003). Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biological Psychiatry*, *54*(5), 504–14.

Poldrack, R.A. (2010). Interpreting developmental changes in neuroimaging signals. *Human Brain Mapping*, *31*(6), 872–8.

Rahko, J., Paakki, J.J., Starck, T., et al. (2010). Functional mapping of dynamic happy and fearful facial expression processing in adolescents. *Brain Imaging and Behavior*, *4*(2), 164–76.

Rangel, A., Camerer, C., Montague, P.R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Neuroscience*, *9*(7), 545–56.

Rex, D.E., Ma, J.Q., Toga, A.W. (2003). The LONI pipeline processing environment. *Neuroimage*, *19*(3), 1033–48.

Romeo, R.D. (2003). Puberty: a period of both organizational and activational effects of steroid hormones on neurobehavioural development. *Journal of Neuroendocrinology*, *15*(12), 1185–92.

Rushworth, M.F., Behrens, T.E., Rudebeck, P.H., Walton, M.E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends in Cognitive Sciences*, *11*(4), 168–76.

Sanz, A., Carrero, P., Pernía, O., Garcia-Segura, L.M. (2008). Pubertal maturation modifies the regulation of insulin-like growth factor-I receptor signaling by estradiol in the rat prefrontal cortex. *Developmental Neurobiology*, *68*(8), 1018–28.

Schulz, K.M., Molenda-Figueira, H.A., Sisk, C.L. (2009). Back to the future: the organizational-activational hypothesis adapted to puberty and adolescence. *Hormones and Behavior*, *55*(5), 597–604.

Shirtcliff, E.A., Dahl, R.E., Pollack, S.D. (2009). Pubertal development: correspondence between hormonal and physical development. *Child Development*, *80*(2), 323–37.

Sisk, C.L., Foster, D.L. (2004). The neural basis of puberty and adolescence. *Nature Neuroscience*, *7*(10), 1040–7.

Somerville, L.H., Jones, R.M., Casey, B.J. (2010). A time of change: behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. *Brain and Cognition*, *72*(1), 124–33.

Sowell, E.R., Trauner, D.A., Gamst, A., Jernigan, T.L. (2002). Development of cortical and subcortical brain structures in childhood and adolescence: a structural MRI study. *Developmental Medicine and Child Neurology*, *44*(1), 4–16.

Spear, L.P. (2004). Adolescent brain development and animal models. *Annals of the New York Academy of Sciences*, *1021*(1), 23–6.

Spear, L.P. (2009). Heightened stress responsivity and emotional reactivity during pubertal maturation: Implications for psychopathology. *Development and Psychopathology*, *21*(01), 87–97.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review : DR*, *28*(1), 78–106.

Steinberg, L. (2010). A behavioral scientist looks at the science of adolescent brain development. *Brain and Cognition*, *72*(1), 160–4.

Steinberg, L., Monahan, K.C. (2007). Age differences in resistance to peer influence. *Developmental Psychology*, *43*, 1531–43.

Tottenham, N., Tanaka, J.W., Leon, A.C., et al. (2009). The nimstim set of facial expressions: judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–9.

Van Leijenhorst, L., Zanolie, K., Van Meel, C.S., Westenberg, P.M., Rombouts, S.A., Crone, E.A. (2010). What motivates the adolescent? Brain regions mediating reward sensitivity across adolescence. *Cerebral Cortex (New York, N.Y. : 1991)*, *20*(1), 61–9.

Wager, T.D., Nichols, T.E. (2003). Optimization of experimental design in fmri: A general framework using a genetic algorithm. *Neuroimage*, *18*(2), 293–309.

Wager, T.D., Davidson, M.L., Hughes, B.L., Lindquist, M.A., Ochsner, K.N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*, *59*(6), 1037–50.

Woods, R.P., Dapretto, M., Sicotte, N.L., Toga, A.W., Mazziotta, J.C. (1999). Creation and use of a talairach-compatible atlas for accurate, automated, nonlinear intersubject registration, and analysis of functional imaging data. *Human Brain Mapping*, *8*(2–3), 73–9.

Downloaded from http://scan.oxfordjournals.org/ at James Cook University on March 15, 2015

Feeding behavior in dopamine-deficient mice

Mark S. Szczypka*, Mark A. Rainey*, Douglas S. Kim*, William A. Alaynick*, Brett T. Marck†, Alvin M. Matsumoto†, and Richard D. Palmiter*‡

*Howard Hughes Medical Institute and Department of Biochemistry, Box 357370, University of Washington, Seattle, WA 98195-7370; and †Gerontology Research Education and Clinical Center, Box 358280, Veterans Administration Medical Center, and Department of Medicine, Seattle, WA 98108

Contributed by Richard D. Palmiter, August 2, 1999

Mice that cannot make dopamine (DA), a condition caused by the selective inactivation of tyrosine hydroxylase in dopaminergic neurons, are born normal but gradually become hypoactive and hypophagic, and die at 3 weeks of age. We characterized the feeding and locomotor responses of these DA-deficient (DA−/−) mice to 3,4-dihyroxy-L-phenylalanine (L-DOPA) to investigate the relationship between brain DA levels and these complex behaviors. Daily administration of L-DOPA to DA−/− mice stimulated locomotor activity that lasted 6 to 9 hr; during that time the mice consumed most of their daily food and water. The minimal dose of L-DOPA that was sufficient to elicit normal feeding behavior in the DA−/− mice also restored their striatal DA to 9.1% of that in the wild-type (WT) mice at 3 hr; then DA content declined to <1% of WT levels by 24 hr. This dose of L-DOPA induced locomotor activity that exceeded that of treated WT mice by 5- to 7-fold, suggesting that DA−/− mice are supersensitive to DA. Unexpectedly, DA−/− mice manifested a second wave of activity 24 to 48 hr after L-DOPA treatment that was equivalent in magnitude to that of WT mice and independent of DA receptor activation. The DA−/− mice approached, sniffed, and chewed food during this second period of activity, but they ate <10% of that required for sustenance. Therefore, DA−/− mice can execute behaviors necessary to seek and ingest food, but they do not eat enough to survive.

Rats treated with 6-hydroxydopamine (6-OHDA) and monkeys treated with *N*-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) have been studied extensively as models of Parkinson's disease because both neurotoxins can be used to destroy the same nigrostriatal dopaminergic neurons that die during the progression of the human disease (1, 2). In animal models and in human Parkinson's disease, behavior is generally unaffected until >80% of the dopamine (DA) is depleted. Hypoactivity, deficits in motor coordination, and failure to eat or drink become evident with more severe DA depletion (3–5). The aphagia and adipsia of rats treated with 6-OHDA can be life-threatening unless they are kept alive by intragastric feeding. With this treatment, the rats gradually become more active, and eventually they eat enough food without assistance to stay alive and gain weight. Compensation for the feeding deficit in lesioned animals is thought to occur through improved efficacy of dopaminergic signaling, because DA levels remain low and a subsequent blockade of tyrosine hydroxylase activity will reinstate aphagia and akinesia (6). Although feeding returns to normal, deficits in feeding responses can be elicited by physiological challenges (4, 5). Neonatal rats with 6-OHDA lesions continue to suckle and can be weaned onto normal chow without major impairments of feeding, despite nearly complete elimination of DA. This suggests that compensatory mechanisms are even more efficacious in young rats (2, 7).

Parkinson's disease differs from the rat model in that the symptoms generally become more severe with age, and, as neurons continue to die, tremor is a prominent feature, but aphagia is generally not a major symptom. Food intake may decline as the disease progresses, but it is not clear whether that is attributable to decreased motivation to eat, difficulty in maneuvering eating utensils, or nonspecific depression. The feeding deficit in the rat model has been attributed to sensorimotor impairment and/or a loss of motivation to eat; however, the mechanisms have not been elucidated. Delivery of DA agonists and antagonists to different sites in the brain has identified specific regions that influence feeding behaviors (8, 9), but they have not indicated where the DA action is essential. Furthermore, the genetic approach of inactivating individual genes encoding DA receptors or transporters has not revealed any striking effects on feeding behavior (10–13). Thus, the specific roles of DA in feeding remain enigmatic.

Gene-targeting techniques were used to inactivate the tyrosine hydroxylase gene in dopaminergic neurons, sparing the production of DA as a precursor for norepinephrine and epinephrine (14). Initial studies in these DA−/− mice revealed that DA is important for control of locomotion, muscle coordination, feeding, and probably other behaviors. The variable extent of 6-OHDA lesions, along with a gradual and variable duration of compensation, makes the study of some behaviors difficult in that model. In contrast, the specificity of genetic lesion (resulting in reproducible and stable behavioral phenotypes) provides favorable background for ascertaining where in the brain, and, ultimately how, DA modulates various behaviors. Prolonged restoration of feeding to physiological levels by pharmacological or gene therapy approaches is rarely described in the 6-OHDA-lesion model, although 3,4-dihyroxy-L-phenylalanine (L-DOPA) and DA agonists can reverse some of the symptoms transiently (15–17). Notable exceptions are studies in which transplantation of dopaminergic tissue revealed that local production of DA in striatum could prevent aphagia induced by 6-OHDA (18, 19). Routine pharmacological treatment induces feeding in DA−/− mice, and the feasibility of long-term rescue of feeding behavior (over a year) by gene therapy has been established (20). Our goal is to map more precisely where DA is required to restore various behaviors, including feeding. A prerequisite to such studies, however, is characterization of behavioral deficits of DA−/− mice.

Materials and Methods

Mice. DA−/− mice were created as described (14). The wild-type (WT) and DA−/− mice in this study were 3- to 6-month-old male and female hybrids of 129/SvEv × C57BL/6J. Because no significant differences between heterozygous and WT mice were observed in any of the parameters measured, they were grouped and designated as WT mice. Mice were housed under standard vivarium conditions and maintained on a 12-hr light/dark cycle with the lights on at 07:00. For routine maintenance of mutants, daily injections of 50 mg/kg L-DOPA were administered to DA−/− mice between 12:00 and 16:00. The genotypes of all mice were assessed by Southern blot analysis.

Activity Measurements. Transparent Plexiglas cages (40 × 20 × 20 cm) were placed on a rack equipped with infrared photobeams (San Diego Instruments) to measure the ambulatory activity of

Abbreviations: DA, dopamine; DA−/−, dopamine-deficient; L-DOPA, 3,4-dihyroxy-L-phenylalanine; 6-OHDA, 6-hydroxydopamine; WT, wild type.

‡To whom correspondence should be addressed. E-mail: palmiter@u.washington.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Downloaded by guest on February 23, 2022

the mice. The number of consecutive beam-breaks that occurred each hour was measured and converted to meters by using the distance between beams (8.8 cm) as a conversion factor.

Neurological and Sensorimotor Examinations. WT (n = 10) and DA−/− (n = 10) mice were subjected to 13 neurological tests and 1 activation test (forced swim) that had previously been used to assess the capacities of 6-OHDA-lesioned rats (21). Catalepsy and biting strength tests were modified to make them suitable for mice. Catalepsy and akinesia tests were performed in the home cage. Mice did not respond to the tactile or noxious stimuli tests designed for rats; these tests were therefore omitted. Tests were performed at 1 and 24 hr after an L-DOPA injection, and mice were scored on a scale of 0 (poor) to 4 (normal).

Food Intake Measurements and Videotaping of Feeding Behavior. Fresh food pellets were placed in the bottom of the cage each day and water was available from 15- or 50-ml bottles equipped with spouts. The amount of food consumed was quantified by weighing the food pellets before and after each test interval and the difference was recorded. Water intake was quantified by weighing bottles in the same fashion. To keep DA−/− mice alive without L-DOPA treatment, Ensure Plus (Abbott Labs) was dispensed in small drops with a syringe and a blunt 30-gauge needle in front of the mouth of hand-held DA−/− mice (n = 3), beginning at postnatal day 21. Mice were routinely maintained on breeder chow (Purina 5015), which contains 11% fat and yields 4.35 kcal/g. For some experiments, we used a standard low-fat chow (Harlan 8604) with 3.93 kcal/g, or a high-fat, high-sucrose, soft palatable chow (Research Diets, D12331) with 5.56 kcal/g. For videotaping of feeding behavior, WT (n = 3) and DA−/− (n = 3) mice were fed the high-fat, high-sucrose, soft, palatable chow for 2 weeks. Food was removed on day 15, and 30 hr later, cage tops were carefully removed and a cup filled with the same type of chow was placed on the bottom of the cage. Reaction of the mice was videotaped for 90 min and the amount of food consumed was determined as described above.

Drug Administration. Drugs were prepared in 10 mM sodium phosphate, 150 mM NaCl, pH = 7.5 (PBS) or 0.9% NaCl. Solutions of L-DOPA (Sigma) at 1.5 mg/ml and the methyl ester of L-DOPA (Research Biochemicals International) at 3 mg/ml were prepared in PBS containing 2.5 mg/ml ascorbic acid (Sigma). Carbidopa (Sigma) was added to L-DOPA solutions at concentrations that allowed for delivery of either 12.5 or 25 mg/kg carbidopa with the standard 50-mg/kg dose of L-DOPA. Amphetamine (Sigma), SCH 23390 (Research Biochemicals International), and haloperidol (McNeil Pharmaceutical) were prepared in 0.9% saline. All drug solutions were delivered via intraperitoneal injection in a volume that did not exceed 1 ml.

D1- and D2-Receptor Antagonist Studies. Mice were placed in activity chambers for at least 12 hr before treatment with 0.9% saline or DA receptor antagonists (SCH 23390 and haloperidol); beam breaks were recorded over 4 hr. The DA−/− mice were treated with 50 mg/kg L-DOPA either 2 or 26 hr before antagonist treatment, such that activity measurements were performed 2 to 6 and 26 to 30 hr after L-DOPA administration.

Monoamine Measurements. The brains of the CO_2-asphyxiated, WT and DA−/− mice were removed and immediately placed in ice-cold PBS for 2 min. Dissections were performed on a glass plate chilled to 4°C, and structures of interest were frozen on dry ice in microcentrifuge tubes and stored at −70°C until analysis. Monoamine measurements were performed as described (22).

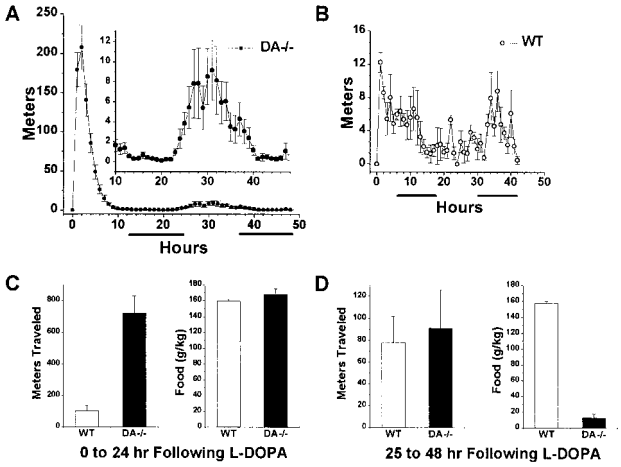


Fig. 1. DA−/− mice exhibit hyperactivity immediately after L-DOPA treatment, and display a second wave of activity that occurs 30 hr after treatment. Ambulatory activity of DA−/− mice (*A*; n = 17] and WT mice (*B*; n = 8). (*C* and *D*) Black bars represent the dark cycle. Total distance traveled and amount of food consumed by DA−/− and WT mice in the first 24-hr period (*C*) and second 24-hr period (*D*) after L-DOPA treatment.

Results

Locomotor Activity of DA−/− Mice After Injection of L-DOPA. Daily administration of 50 mg/kg L-DOPA into DA−/− mice induced locomotor activity that greatly exceeded that of L-DOPA-treated WT mice for 6 to 9 hr; by 24 hr, however, the DA−/− mice had become hypoactive (Fig. 1*A*). Because the DA−/− mice remained hyperactive for an entire day when injected every 4 hr with 100 mg/kg L-DOPA, and they traveled up to 3000 meters/day (data not shown), the decline in activity after the first few hours was not caused by fatigue. The overall distance traveled by DA−/− mice during the first 24 hr after treatment was 7-fold greater than that of WT mice (Fig. 1*C*). An unexpected second wave of activity occurred between 24 and 40 hr after the previous L-DOPA treatment (Fig. 1*A*, *Inset*). The mean distances traveled by DA−/− and WT mice between 24 and 48 hr after an L-DOPA injection were equivalent (Fig. 1*D*); however, during this period of activity DA−/− mice typically displayed a hunched appearance, walked with an awkward gait, and performed poorly on rotarod and pole tests (23).

The distance traveled by DA−/− mice increased marginally as the dose of L-DOPA was increased to 20 mg/kg, then steadily increased to reach a plateau between 40 and 100 mg/kg L-DOPA (Fig. 2*A*). None of these dosages of L-DOPA had any effect on WT mice.

Neurological and Sensorimotor Examination of WT and DA−/− Mice. Mice performed well on many of the neurological tests (Table 1). DA−/− mice scored 486 (1 hr after L-DOPA), but only 379 (24 hr after L-DOPA) out of a possible 520 points, compared to 495 for WT mice. DA−/− mice tested 24 hr after L-DOPA treatment displayed greater one- and four-limb akinesia and forelimb catalepsy than did WT mice (Table 1). When tested 1 hr after L-DOPA treatment, the DA−/− mice were deficient in the paw-pinch assay. The performance of the DA−/− mice in the forced-swim test 24 hr after L-DOPA treatment was equivalent to that of the WT mice.

Food Consumption of DA−/− Mice in Response to L-DOPA. DA−/− mice can be maintained for at least 1.5 yr with daily administration of 50 mg/kg L-DOPA if they are fed breeder chow (4.35 kcal/g); standard chow (3.93 kcal/g), however, is inadequate for survival. With treated daily with L-DOPA injections at 25 mg/kg,

Downloaded by guest on February 23, 2022

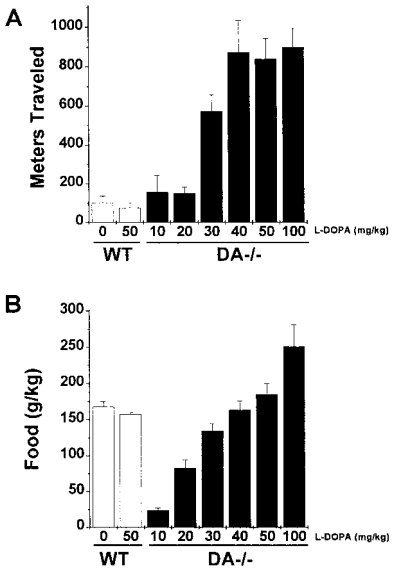


Fig. 2. Low doses of L-DOPA cause hyperactivity in DA−/− mice. Individually housed DA−/− mice during the 9-hr period after administration of 10, 20, 30, 40, or 50 mg/kg or 100 mg/kg L-DOPA methyl ester were used to generate a dose-response curve. Various doses of L-DOPA were administered 24 hr after their last L-DOPA injection. (*A*) Ambulatory activity of WT ($n = 8$) and DA−/− mice ($n = 8$). (*B*) Food consumed by WT ($n = 8$) and DA−/− mice ($n = 8$).

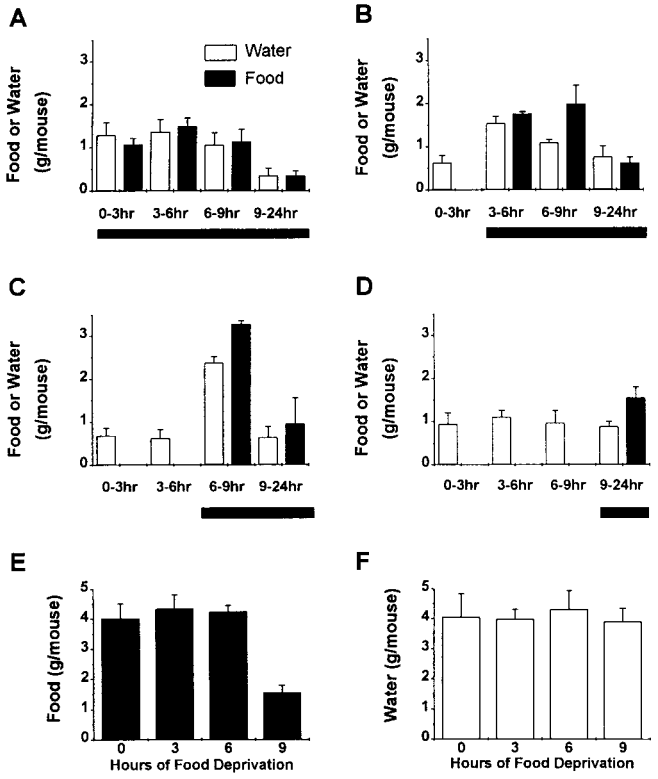


Fig. 3. Food deprivation identifies the minimal amount of DA required to sustain food consumption in DA−/− mice. Mice ($n = 15$) housed in groups of three were treated with 50 mg/kg L-DOPA, returned to their home cages, and food was added after 3, 6, or 9 hr of restriction. Water was available ad libitum during the food restriction studies. Amount of food and water consumed after 0 (*A*), 3 (*B*), 6 (*C*), or 9 (*D*) hr of food deprivation is shown. Black horizontal lines represent the time when food was present in the cage. (*E*) Total amount of food consumed when mice were food-deprived after L-DOPA treatment. (*F*) Water consumed by food-deprived DA−/− mice.

the DA−/− mice lost weight and would have succumbed if this regimen had continued.

DA−/− mice housed together consumed about 30% of their total daily food and water during each of the first three 3-hr periods, and the remaining 10% was consumed during the next 15 hr (Fig. 3*A*). Water and food consumption by the two groups were comparable during the 24 hr. The total amounts of food consumed by DA−/− and WT mice during the first 24 hr after L-DOPA treatment were similar; however, during the next 24 hr, the DA−/− mice ate only 10% as much food as did the WT mice (Fig. 1 *C* and *D*). Food intake was proportional to the L-DOPA dose up to 100 mg/kg (Fig. 2*B*).

In another experiment, DA−/− mice were injected with L-DOPA but food was unavailable until 3, 6, or 9 hr later; water was available all the time. When food was restored 3 or 6 hr after the L-DOPA injection, the DA−/− mice compensated by eating more during the next 3-hr period (Fig. 3 *B* and *C*); by 24 hr, they had consumed the same amount of food as under ad libitum conditions (Fig. 3*E*). However, if food was restored at 9 hr, the DA−/− mice ate only 35% of the daily requirement (Fig. 3 *D* and *E*), but they consumed normal amounts of water even when food consumption fell below normal (Fig. 3 *E* and *F*). Although food and water consumption normally occurred in parallel (Fig. 3*A*), they can be dissociated (Fig. 3 *B–D*).

Table 1. Neurological testing of WT and DA−/− mice

	Test score		
	WT	DA−/− mice	
Test	24 hr	1 hr	24 hr
Screen paw placement	40	40	40
One-limb akinesia	39	40	14***
Four-limb akinesia	40	40	9***
Turning on a tilted screen	35	29	7**
Forelimb catalepsy	40	39	14***
Visual placing	36	39	32
Orientation to paw pinch	40	33*	37
Auditory response	40	40	40
Vestibular response	40	40	40
Splay posture	40	40	40
Biting strength	25	26	26
Righting reflex	40	40	40
Hind-limb extensor rigidity	40	40	40
Total	495	486	379

Neurological tests were performed on WT and DA−/− mice 1 and 24 hr after L-DOPA treatment. No differences were observed in WT mice at 1 and 24 hr. Maximum possible score for each test was 40 (4 × 10 mice). Significance was assessed by standard ANOVA. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ compared with WT mice.

The Feeding Behavior of DA−/− Mice. Hand-held, young DA−/− pups that had never been injected with L-DOPA would lick and swallow small drops of a liquid diet placed by their mouth. With frequent feeding by this method, the mice ate enough to maintain body weight for a week without L-DOPA treatment; however, if left with their mother, they would have died.

When presented with a palatable, high-sucrose, high-fat diet after a 30-hr fast, the adult DA−/− and WT mice approached, sniffed, and began to consume the food; but consumption by the DA−/− mice was short-lived compared with that of the WT mice. After 90 min, most of the DA−/− mice were inactive or sleeping, whereas WT mice were still moving about the cage and occasionally eating. The amount of food eaten by the DA−/−

Downloaded by guest on February 23, 2022

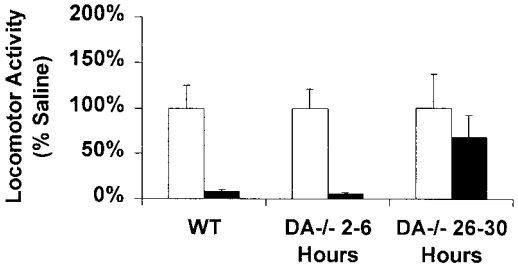


Fig. 4. DA receptor antagonists do not block the second wave of locomotion displayed by DA−/− mice. Ambulatory activity of WT and DA−/− mice treated with 0.9% saline (open bars) or a mixture of 0.1 mg/kg SCH 23390 and 2 mg/kg haloperidol (closed bars) shown as percent of saline-elicited activity. Meters traveled ± SEM: WT saline = 24.6 ± 6.2; DA−/− saline (2–6 hr) = 281.1 ± 59.3; and DA−/− saline (26–30 hr) = 17.9 ± 6.7.

mice was only 25% that of the WT mice during this 90-min period.

The Second Wave of Activity. The second wave of activity may have been caused by the release of residual DA. To test this possibility, a mixture of D1- and D2-receptor antagonists (SCH 23390 and haloperidol, respectively) was administered during the second wave of activity and locomotion was monitored. The DA-receptor antagonists reduced locomotion by more than 90% in WT and DA−/− mice during the first wave of activity (Fig. 4). The same mixture of antagonists had no significant effect on the locomotion of DA−/− mice during the second wave of activity (Fig. 4).

Amphetamine stimulates locomotion in rodents by inducing the release of DA from dopaminergic terminals (24, 25). Administration of 5 mg/kg amphetamine to DA−/− mice 19 hr after their last treatment with L-DOPA stimulated ambulatory activity. However, the response of DA−/− mice was completed within 1 hr, compared with 2 hr for WT mice (Fig. 5*A*). The WT mice responded similarly to a second dose of amphetamine given 2 hr after the first dose, whereas the DA−/− mice were completely unresponsive to a second dose (Fig. 5*A*). The locomotor response of DA−/− mice to amphetamine was completely blocked by a combination of D1- and D2-receptor antagonists (data not shown). Fig. 5*B* demonstrates that a second wave of activity occurred on schedule even when residual DA was depleted by amphetamine.

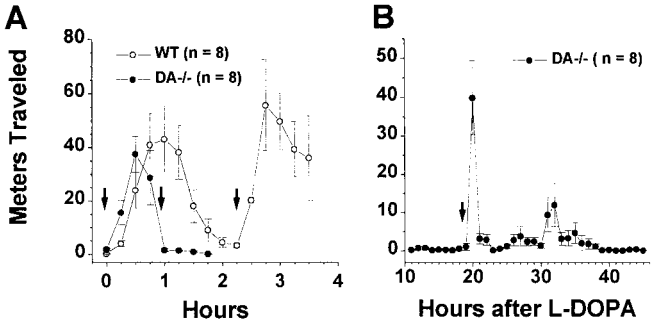


Fig. 5. Amphetamine induces locomotor behavior in DA−/− mice, but amphetamine treatment does not block the second wave of activity displayed by DA −/− mice. (*A*) Ambulatory activity of WT and DA−/− mice given two consecutive injections of 5 mg/kg amphetamine. Arrows indicate times of amphetamine administration. (*B*) Amphetamine (5 mg/kg) was administered 19 hr after the L-DOPA injection and activity was monitored for an additional 36 hr. Arrow indicates time of amphetamine injection.

Table 2. Catecholamine content in WT and DA−/− mice

Group	Brain region	*n*	Time after treatment, hr	L-DOPA, mg/kg	DA, ng/mg protein
WT	Striatum	7	0	Naïve	108.05 ± 12.02
		10	3	50	91.14 ± 7.98
		5	30	50	89.14 ± 10.15
DA−/−		5	0	Naïve	0.48 ± 0.05
		10	3	50	8.31 ± 1.46*,**
		4	9	50	3.47 ± 0.54*
		5	30	50	0.50 ± 0.05*
		4	3	100†	25.79 ± 6.04*,**
		8	3	50 + CD	101.86 ± 18.77**
WT	Thalamus	10	3	50	12.29 ± 1.87
		5	30	50	13.29 ± 2.52
DA−/−		12	3	50	3.13 ± 0.55*
		5	30	50	0.98 ± 0.16*
		8	3	50 + CD	39.28 ± 2.48*,**
WT	Midbrain	7	0	Naïve	2.89 ± 0.35
		9	3	50	2.37 ± 1.38
		6	30	50	2.15 ± 0.87
DA−/−		5	0	Naïve	0.47 ± 0.07*
		10	3	50	1.61 ± 0.29
		4	9	50	0.81 ± 0.05
		5	30	50	0.63 ± 0.13*
		4	3	100†	1.17 ± 0.16
		4	3	50 + CD	8.71 ± 1.03*,**

WT and DA−/− mice were injected with L-DOPA, L-DOPA methy ester (†), or L-DOPA + 25 mg/kg carbidopa (CD), and animals were sacrificed at 3, 6, 9, or 30 hr after treatment. Naïve mice did not receive injection of L-DOPA and were sacrificed when 17 to 20 days old. Brains from all mice were removed and dissected and catecholamine concentrations were measured with HPLC and electrochemical detection. Significance was assessed by standard two-tailed *t*-test analysis. *, $P < 0.05$ compared with WT mice; **, $P < 0.05$ compared with naïve DA−/− mice.

Brain DA Content Is Only Partially Restored by L-DOPA. At the peak of activity, 3 hr after the standard dose of 50 mg/kg L-DOPA, the highest DA content was found in the striatum, but it was only 9.1% of the WT levels (Table 2). At 9 hr, when there was insufficient DA to support normal food consumption (Fig. 3*E*), DA content was 3.8% of normal; at 30 hr, DA levels were <1% of normal (Table 2). The ratio of DA to norepinephrine in the striatum of naïve DA−/− mice never injected with L-DOPA was 0.08, not significantly different from the ratio obtained at 30 hr (0.1). This value is also similar to the ratio observed in brain regions that contain few dopaminergic projections (e.g., cerebellum) and several peripheral tissues (22). Administration of 100 mg/kg of the L-DOPA methyl ester raised brain DA content in the striatum (28.2% of normal) but not in the midbrain (Table 2). Norepinephrine and serotonin levels were normal in DA−/− mice 3 and 30 hr after L-DOPA treatment (data not shown).

Carbidopa Elevates Brain DA Concentrations and Induces Stereotypy. To achieve higher levels of L-DOPA in the brain, peripheral L-aromatic amino acid decarboxylase activity was inhibited by coinjecting carbidopa. Administration of L-DOPA and carbidopa restored DA content within the striatum to normal levels and to more than normal levels in the midbrain (Table 2). In addition, the concentration of the primary DA metabolite, dihydroxyphenylacetic acid (DOPAC), was elevated in all brain

Downloaded by guest on February 23, 2022

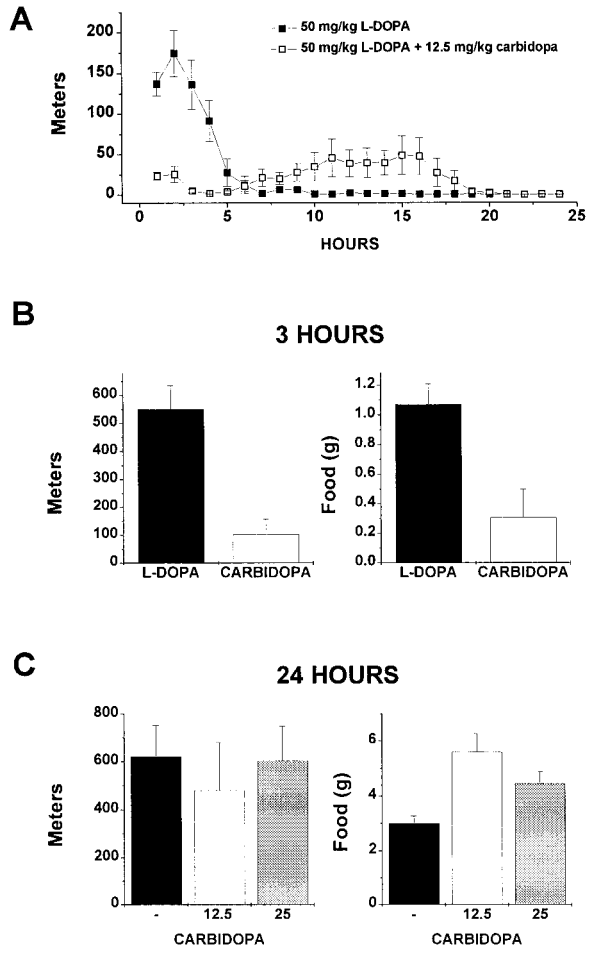


Fig. 6. Simultaneous carbidopa and L-DOPA treatment induces stereotypy in DA−/− mice. (*A*) Ambulatory activity of DA −/− mice ($n = 8$) treated with 50 mg/kg L-DOPA (■) or a mixture of 50 mg/kg L-DOPA and 12.5 mg/kg carbidopa (□) parse by 1-hr intervals. (*B*) Total meters traveled and food consumed by DA−/− mice in the first 3 hr after injection of 50 mg/kg L-DOPA alone (black bars) or 50 mg/kg L-DOPA and 12.5 mg/kg carbidopa (open bars). (*C*) Total meters traveled and food consumed by DA−/− mice in 24 hr when treated with 50 mg/kg L-DOPA alone (black bars), a mixture of 50 mg/kg L-DOPA and 12.5 mg/kg (open bars), or 50 mg/kg L-DOPA and 25 mg/kg carbidopa (gray bars).

regions, suggesting that DA turnover was elevated (data not shown). When carbidopa at 12.5 or 25 mg/kg was given with the standard 50 mg/kg dose of L-DOPA, it produced a biphasic activity profile, with a lull in ambulatory activity between 1 and 5 hr (Fig. 6*A*). During this lull, DA−/− mice typically sat upright with their paws together and engaged in intense stereotypic behavior that included licking and chewing of paws and occasionally self-mutilation. This treatment produced less locomotor activity and feeding during the 3 hr after treatment than that achieved by L-DOPA alone (Fig. 6*B*), but activity resumed and persisted much longer, such that during the 24 hr, these mice traveled a distance similar to that when they were given L-DOPA alone. The mice also ate more food with either dose of carbidopa (Fig. 6*C*). WT mice treated with L-DOPA and either dose of carbidopa became inactive and did not exhibit stereotypy.

Discussion

There is extensive literature implicating DA in movement, sensorimotor function, and feeding. DA−/− mice are unable to synthesize DA in dopaminergic terminals without pharmacological intervention or gene therapy, and consequently, when depleted of DA, they display defects in many of these activities. When DA−/− mice are treated with L-DOPA, the concentration of DA restored in the brain is only a fraction of WT levels (Table 2), but they manifest locomotor activity that reaches 200 meters/hr, about 20 times that of WT mice (Fig. 1*A*). Food consumption is normal with this routine L-DOPA treatment (Fig. 1*C*). These results demonstrate that DA−/− mice are supersensitive to brain DA. This supersensitivity to DA does not appear to be caused by differences in the number of DA receptors or in the DA transporter, as measured by ligand binding, suggesting that it resides in more efficient coupling of DA receptors to the signal transduction (unpublished observations).

Adult 6-OHDA-lesioned rats fail to move, eat, or drink for several days after surgery, but with assistance, the rats will progress through several stages of behavioral recovery and eventually regain the ability to perform these behaviors (26). In the early stages of recovery, a phenomenon called "paradoxical kinesia" can be elicited in which activity and feeding can be restored for a short time by presentation of novel stimuli (21, 27). DA−/− mice resemble rats in this early stage; they become hypoactive, aphagic, and adipsic, and "paradoxical kinesia" can be induced 24 hr after L-DOPA treatment. Even without stimulation, DA−/− mice undergo a DA-independent wave of motor activity 24–48 hr after L-DOPA treatment (Figs. 1 and 4). During this second wave of activity, the DA−/− mice approach, sniff, chew, and swallow food, but they consume only a fraction of that needed for sustenance. In the rat 6-OHDA model, there is gradual compensation for the feeding deficit when the lesions are performed in adults, and no deficits are observed after neonatal lesions; in both cases, brain DA levels are reported as $<$ 2% of normal. Although increases in serotonin have been observed (28, 29), normal feeding continues to depend on DA; i.e., a second 6-OHDA treatment or inhibition of tyrosine hydroxylase activity reinstates aphagia (6). We observe insufficient feeding in DA−/− mice when DA falls below 3.8% of normal (9 hr after L-DOPA administration), serotonin levels are unaltered, and compensation is not observed even after a year of daily L-DOPA treatments.

Responses of DA−/− mice in 9 of 13 tests that involve external stimuli, as well as the forced-swim test, were normal and unaffected by L-DOPA treatment. However, their responses in 4 tests that involve voluntary or coordinated movement were impaired 24 hr after L-DOPA but were restored 1 hr after L-DOPA treatment (Table 1). This finding is in agreement with the inability of DA−/− mice to perform other motor tasks that require coordinated movement (pole and rotarod test) when DA is depleted (14, 20). Overall, DA −/− mice performed at 76.6% (1 hr after L-DOPA) and 98.2% (24 hr after L-DOPA) of WT mice on these tests. DA−/− mice ambulate normal distances during the second 24 hr period after L-DOPA treatment without consuming the normal amount of food. However, they ate normally for a short time when presented with palatable food during that period. They could also be kept alive for over a week without L-DOPA treatment by hand-feeding them a liquid diet. These observations indicate that DA−/− mice can execute behaviors required for seeking and ingesting food; however, without L-DOPA treatment they will die of starvation even with easy access to palatable food. Likewise, DA−/− pups with a nursing mother will not suckle enough to sustain themselves.

It is noteworthy that adult rats with severe 6-OHDA lesions were maintained by gastric feeding rather than pharmacologically. In that model, dopaminergic terminals are almost completely destroyed and those that remain probably manifest maximal activity; hence DA release cannot be enhanced by administering L-DOPA. We have found it difficult to restore normal feeding (or other behaviors) of DA−/− mice with

Downloaded by guest on February 23, 2022

direct-acting DA agonists and the same may be true for lesioned rats. Thus, activity-dependent release of DA from dopaminergic terminals (perhaps along with other neuromodulators) may be necessary for normal feeding.

An optimal level of dopaminergic signaling may be required for the mice to engage in normal behaviors. Peripherally administered DA agonists or antagonists inhibit feeding in both WT and partially lesioned rats, suggesting that acute elevation or dampening of DA signaling adversely affects rat feeding behavior (8, 30). DA−/− mice also have an optimal range of brain DA levels (3.5–25% of normal) that can support normal feeding. When DA falls below that range, the mice do not consume adequate amounts of food; when DA rises above that range, intense stereotypic behaviors interfere with feeding. Elevation of brain DA concentrations to WT levels causes DA−/− mice to engage in intense stereotypy and forego eating for several hours; but when DA levels begin to decline, they ambulate and feed as when injected with L-DOPA alone (Fig. 6 *A–C*). The maximal locomotor response observed with this treatment does not reach the magnitude obtained with L-DOPA treatment alone, but activity and feeding are prolonged.

What neuromodulator(s) induce the second wave of locomotor activity? We suspect that the second wave of activity is the consequence of a circadian input. Considering that dopaminergic neurons are intact in this model, the simplest suggestion is that these neurons normally liberate neuromodulators (in addition to DA) that contribute to voluntary locomotion. In the absence of DA, these other neuromodulators may continue to stimulate locomotion in response to circadian inputs. Neuromodulators/neurotransmitters, such as cholecystokinin, neurotensin, ATP, adenosine, or glutamate are candidates (31–35). Alternatively, the circadian activity may be completely independent of inputs from dopaminergic neurons.

The DA−/− mice reveal that DA is essential for normal feeding. The same conclusion was reached many years ago by several investigators studying rats with 6-OHDA lesions; however, the insult produced by destroying neurons greatly exceeds that resulting from the removal of a single enzyme. Thus, it is remarkable that the 6-OHDA lesions and genetic model are so similar, despite the fact that dopaminergic neurons produce many neuromodulators in addition to DA.

We thank Dr. Ed Stricker for many helpful suggestions during the preparation of this manuscript. M.S.S. was supported by National Institutes of Health Grant HD-08121. D.S.K. is in the Molecular and Cellular Biology Program at the University of Washington and the Fred Hutchinson Cancer Research Center and was supported by a National Science Foundation Graduate Research Fellowship. This work was also supported in part by National Institutes of Health Grant HD-09172.

1. Kopin, I. J. & Markey, S. P. (1988) *Annu. Rev. Neurosci.* **11,** 81–96.
2. Zigmond, M. J. & Stricker, E. M. (1989) *Int. Rev. Neurobiol.* **31,** 1–79.
3. Ungerstedt, U. (1971) *Acta Physiol. Scand. Suppl.* **367,** 95–122.
4. Zigmond, M. J. & Stricker, E. M. (1972) *Science* **177,** 1211–1214.
5. Marshall, J. F. & Teitelbaum, P. (1973) *Brain Res.* **55,** 229–233.
6. Zigmond, M. J., Abercrombie, E. D., Berger, T. W., Grace, A. A. & Stricker, E. M. (1990) *Trends Neurosci.* **13,** 290–296.
7. Smith, R. D., Cooper, B. R. & Breese, G. R. (1973) *J. Pharmacol. Exp. Ther.* **185,** 609–619.
8. Heffner, T. G., Zigmond, M. J. & Stricker, E. M. (1977) *J. Pharmacol. Exp. Ther.* **201,** 386–399.
9. Maldonado-Irizarry, C. S., Swanson, C. J. & Kelley, A. E. (1995) *J. Neurosci.* **15,** 6779–6788.
10. Drago, J., Gerfen, C. R., Lachowicz, J. E., Steiner, H., Hollon, T. R., Love, P. E., Ooi, G. T., Grinberg, A., Lee, E. J., Huang, S. P., *et al.* (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12564–12568.
11. Xu, M., Moratalla, R., Gold, L. H., Hiroi, N., Koob, G. F., Graybiel, A. M. & Tonegawa, S. (1994) *Cell* **79,** 729–742.
12. Baik, J. H., Picetti, R., Saiardi, A., Thiriet, G., Dierich, A., Depaulis, A., Le Meur, M. & Borrelli, E. (1995) *Nature (London)* **377,** 424–428.
13. Giros, B., Jaber, M., Jones, S. R., Wightman, R. M. & Caron, M. G. (1996) *Nature (London)* **379,** 606–612.
14. Zhou, Q. Y. & Palmiter, R. D. (1995) *Cell* **83,** 1197–1209.
15. Ljungberg, T. & Ungerstedt, U. (1976) *Physiol. Behav.* **16,** 277–283.
16. Marshall, J. F. & Ungerstedt, U. (1976) *Physiol. Behav.* **17,** 817–822.
17. Ungerstedt, U. & Ljungberg, T. (1977) *Adv. Biochem. Psychopharmacol.* **16,** 193–199.
18. Schwarz, S. S. & Freed, W. J. (1987) *Exp. Brain Res.* **65,** 449–454.
19. Rogers, D. C. & Dunnett, S. B. (1990) *Prog. Brain. Res.* **82,** 489–492.
20. Szczypka, M. S., Mandel, R. J., Donahue, B. A., Snyder, R. O., Leff, S. E. & Palmiter, R. D. (1999) *Neuron* **22,** 167–178.
21. Marshall, J. F., Levitan, D. & Stricker, E. M. (1976) *J. Comp. Physiol. Psychol.* **90,** 536–546.
22. Thomas, S. A., Marck, B. T., Palmiter, R. D. & Matsumoto, A. M. (1998) *J. Neurochem.* **70,** 2468–2476.
23. Matsuura, K., Kabuto, H., Makino, H. & Ogawa, N. (1997) *J. Neurosci. Methods* **73,** 45–48.
24. Jones, S. R., Gainetdinov, R. R., Wightman, R. M. & Caron, M. G. (1998) *J. Neurosci.* **18,** 1979–1986.
25. Fibiger, H. C., Fibiger, H. P. & Zis, A. P. (1973) *Br. J. Pharmacol.* **47,** 683–692.
26. Marshall, J. F. & Teitelbaum, P. (1974) *J. Comp. Physiol. Psychol.* **86,** 375–395.
27. Keefe, K. A., Salamone, J. D., Zigmond, M. J. & Stricker, E. M. (1989) *Arch. Neurol.* **46,** 1070–1075.
28. Stachowiak, M. K., Bruno, J. P., Snyder, A. M., Stricker, E. M. & Zigmond, M. J. (1984) *Brain Res.* **291,** 164–167.
29. Blue, M. E. & Molliver, M. E. (1987) *Brain Res.* **429,** 255–269.
30. Zigmond, M. J., Heffner, T. G. & Stricker, E. M. (1980) *Prog. Neuropsychopharmacol.* **4,** 351–362.
31. Ferré, S., Herrera-Marschitz, M., Grabowska-Andén, M., Ungerstedt, U., Casas, M. & Andén, N. E. (1991) *Eur. J. Pharmacol.* **192,** 25–30.
32. Kiyama, H., McGowan, E. M. & Emson, P. C. (1991) *Brain Res. Mol. Brain Res.* **9,** 87–93.
33. Augood, S. J., Westmore, K. & Emson, P. C. (1997) *Neuroscience* **76,** 763–774.
34. Fuxe, K., Ferré, S., Zoli, M. & Agnati, L. F. (1998) *Brain Res. Brain Res. Rev.* **26,** 258–273.
35. Sulzer, D., Joyce, M. P., Lin, L., Geldwert, D., Haber, S. N., Hattori, T. & Rayport, S. (1998) *J. Neurosci.* **18,** 4588–4602.

Downloaded by guest on February 23, 2022

NEUROSCIENCE

Coupled electrophysiological, hemodynamic, and cerebrospinal fluid oscillations in human sleep

Nina E. Fultz[1,2], Giorgio Bonmassar[2,3], Kawin Setsompop[2,3], Robert A. Stickgold[4,5], Bruce R. Rosen[2,3], Jonathan R. Polimeni[2,3], Laura D. Lewis[1,2]*

Sleep is essential for both cognition and maintenance of healthy brain function. Slow waves in neural activity contribute to memory consolidation, whereas cerebrospinal fluid (CSF) clears metabolic waste products from the brain. Whether these two processes are related is not known. We used accelerated neuroimaging to measure physiological and neural dynamics in the human brain. We discovered a coherent pattern of oscillating electrophysiological, hemodynamic, and CSF dynamics that appears during non–rapid eye movement sleep. Neural slow waves are followed by hemodynamic oscillations, which in turn are coupled to CSF flow. These results demonstrate that the sleeping brain exhibits waves of CSF flow on a macroscopic scale, and these CSF dynamics are interlinked with neural and hemodynamic rhythms.

Sleep is crucial for both high-level cognitive processing and also basic maintenance and restoration of physiological function. During human non–rapid eye movement (NREM) sleep, the electroencephalogram (EEG) exhibits low-frequency (<4 Hz) oscillatory dynamics that support memory and neural computation (*1*–*8*). In addition, functional magnetic resonance imaging (fMRI) studies measuring blood oxygen level–dependent (BOLD) signals have demonstrated widespread hemodynamic alterations during NREM sleep (*9*–*15*). Sleep is also associated with increased interstitial fluid volume and clearance of metabolic waste products into the CSF (*16*), and clearance is stronger in sleep with more low-frequency EEG oscillations (*17*). Why these diverse physiological processes co-occur in this state of low arousal is not known. In particular, it remains unclear how CSF dynamics change during sleep and how they relate to the major changes in neural activity and hemodynamics.

We simultaneously measured BOLD fMRI dynamics, EEG, and CSF flow during human sleep. To achieve high–temporal-resolution imaging, we acquired fMRI data at fast rates [repetition time (TR) < 400 ms]. While fMRI is often used to detect local oxygenation changes, fast acquisition paradigms also enable detection of fluid inflow: Fresh fluid arriving at the edge of the imaging volume has high signal intensity because it has not yet experienced radiofrequency pulses (fig. S1). By placing the boundary edge of the imaging volume at the fourth ventricle (Fig. 1A), CSF flow into the brain was detected as increased signal in the lower slices (Fig. 1B), allowing us to measure dynamics of CSF flow simultaneously with BOLD fMRI. We combined this imaging with simultaneous EEG (n = 13 participants) and identified continuous segments of clear stable wake or NREM sleep with low motion (fig. S2 and materials and methods) to enable analysis of continuous low-frequency dynamics.

We first investigated whether sleep was associated with distinct CSF flow dynamics (Fig. 1, C to E). During wakefulness, the CSF signal exhibited a small-amplitude rhythm synchronized to the respiratory signal at ~0.25 Hz (Fig. 1, E and G), consistent with previous studies (*18*, *19*). By contrast, during NREM sleep, we observed a large oscillation in the CSF signal at 0.05 Hz (Fig. 1, E and G). We analyzed this CSF signal across all sleep segments, confirming that identified sleep segments exhibited low-frequency EEG signatures of NREM sleep (Fig. 1F). We found a 5.52-dB increase in the CSF signal peaking at 0.05 Hz during sleep (Fig. 1, G and H, and fig. S2) [95% confidence interval (CI) = (2.33, 7.67); P = 0.003, signed-rank test], suggesting that large waves of CSF inflow occur approximately every 20 s. We additionally analyzed nearby non-CSF regions of interest (ROIs) with matched slice positioning and saw no such effect [change = −0.03 dB, CI = (−2.7, 1.3); P = 0.003 for difference, signed-rank test], suggesting that this sleep-associated pattern was specifically driven by physiological signals in the ventricle (Fig. 1I).

Because inflow signals are caused by fluid flowing into the acquisition volume, CSF flow signals should be brightest in edge slices and decay as fluid passes into central slices (Fig. 2, A and B, and fig. S1). We indeed observed a gradient of signal amplitudes across the slices (Fig. 2, C and D). Some large inflow events exhibited equally bright amplitudes across the lower slices (Fig. 2D), suggesting the CSF flow velocity had exceeded the imaging critical velocity (11.4 mm/s for slice 2). Together, these results identified a large-amplitude pulsatile flow of CSF at 0.05 Hz that appears during NREM sleep.

We next examined whether these slow macroscopic CSF oscillations were linked to hemodynamic signals. Previous studies have proposed microscopic arterial pulsation, corresponding to the ~1-Hz cardiac cycle, as a mechanism for driving interstitial fluid flow (*20*–*23*). To test what might generate the much slower macroscopic CSF rhythm we observed, we analyzed the changes in BOLD signal during sleep. We observed an increase in BOLD signal amplitude in the cortical gray-matter fMRI signal during sleep, as compared with wakefulness (Fig. 3A versus B and C) [mean = 3.28 dB; CI = (0.09, 6.54); P = 0.032, signed-rank test], consistent with previous reports of low-frequency BOLD fluctuations during sleep (*9*, *10*, *24*). Furthermore, the CSF signal was tightly temporally coupled to the cortical gray-matter BOLD oscillation during sleep (Fig. 3, A and B), exhibiting a strong anticorrelation (fig. S3) (maximal r = −0.48 at lag 2 s, P < 0.001, shuffling).

This anticorrelation suggested a possible alternation of blood flow and CSF flow during sleep. We hypothesized that the BOLD oscillations corresponded to an oscillation in cerebral blood volume and that, because of constant intracranial volume, more CSF flows into the head when less volume is occupied by the blood (*25*, *26*). This hypothesis predicts that the CSF signal should approximately match the negative derivative of the BOLD oscillation, after setting negative values to zero (materials and methods). Consistent with this hypothesis, the CSF time series and the thresholded derivative BOLD signals were strongly correlated (Fig. 3, D and E) (maximal r = 0.59 at lag −1.8 s; P < 0.001, shuffling).

We next examined whether neural activity was linked to these coupled hemodynamic and CSF oscillations during sleep. In conventional fMRI, the BOLD response is elicited by neural activity, which drives flow of oxygen-rich blood (*27*, *28*), and EEG oscillations are associated with hemodynamic signals (*29*–*32*). We therefore hypothesized that the large, slow-delta (0.2 to 4 Hz) electrophysiologic oscillations characteristic of NREM sleep could be coupled to oscillations in blood volume and, in turn, displacement effects on CSF flow. We analyzed the instantaneous amplitude of slow-delta EEG relative to the peak of the CSF waves and found that neural, BOLD, and CSF waves were coupled (Fig. 4, A to C, and fig. S4). The neural waves preceded the CSF waves, with a peak in slow-delta EEG occurring 6.4 s before the CSF peak (peak amplitude = 21%, P < 0.001, shuffling). We calculated the best-fit impulse response between the EEG and CSF (Fig. 4D)

[1]Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA. [2]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA 02129, USA. [3]Department of Radiology, Harvard Medical School, Boston, MA 02115, USA. [4]Department of Psychiatry, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA. [5]Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA.
*Corresponding author. Email: ldlewis@bu.edu

Downloaded from http://science.sciencemag.org/ on November 5, 2019

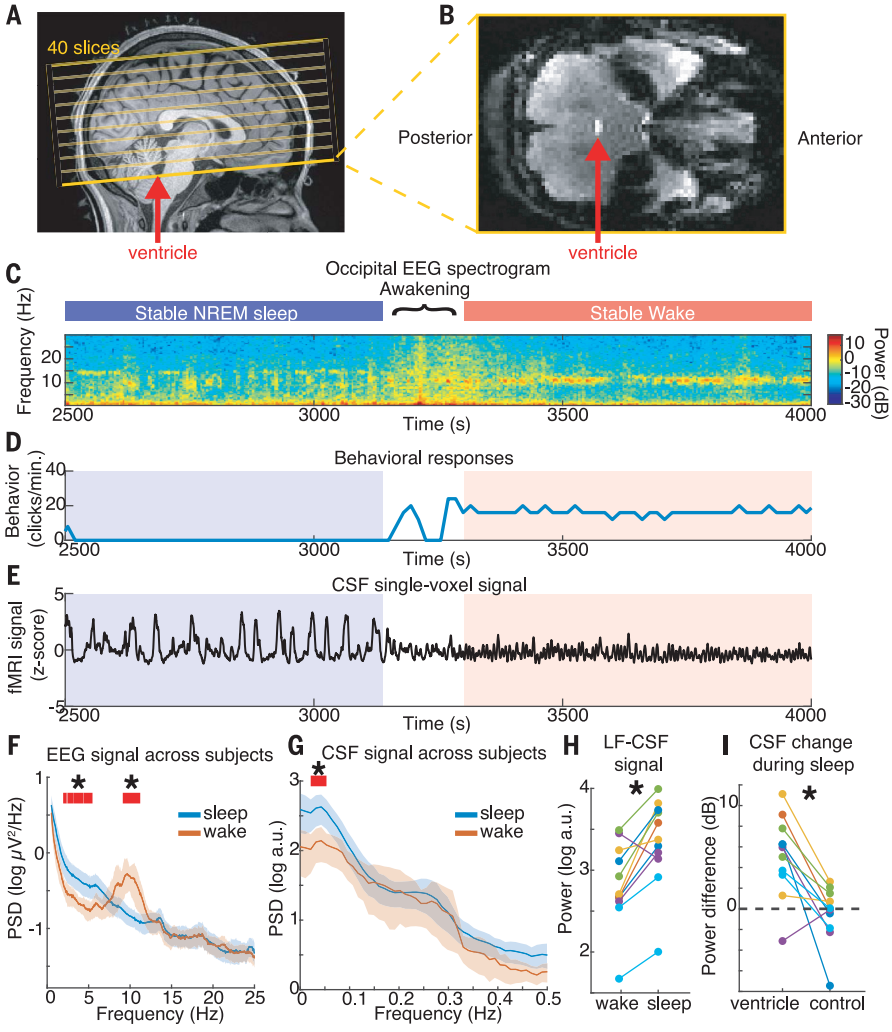


Fig. 1. Large oscillations in CSF signals appear in the fourth ventricle during sleep. (**A**) Example scan positioning. Thick yellow line: position of the functional image relative to the anatomy. The bottom edge intersects with the fourth ventricle (red arrow), allowing CSF inflow to be measured. A subset of the 40 acquired slices are displayed. (**B**) Example functional image from the bottom slice. Inflow through the ventricle is detected as a bright signal (red arrow). (**C**) EEG spectrogram from this individual shows long periods of NREM sleep and wake (~10 Hz occipital alpha). (**D**) Behavioral responses from this individual. (**E**) Time series of a single CSF voxel (smoothed with 10-TR kernel) shows large, slow dynamics in sleep that subside during wakefulness. (**F**) Mean power spectral density (PSD) of occipital EEG confirms slow-delta power in sleep, as opposed to high alpha power in wake (n = 13 participants sleep; 11 participants wake). (**G**) PSD of CSF signal shows increased 0.05 Hz power during sleep (n = 13 participants sleep; 11 participants wake). Shaded regions denote 95% CIs; red lines and asterisks mark non-overlapping CIs. (**H**) Low-frequency (LF, 0 to 0.1 Hz) CSF power increased during sleep (n = 11 participants for pairwise comparison). a.u., arbitrary units. (**I**) This sleep-selective power increase was specific to the ventricle ROI and not observed in a neighboring size-matched control ROI (n = 11 participants).

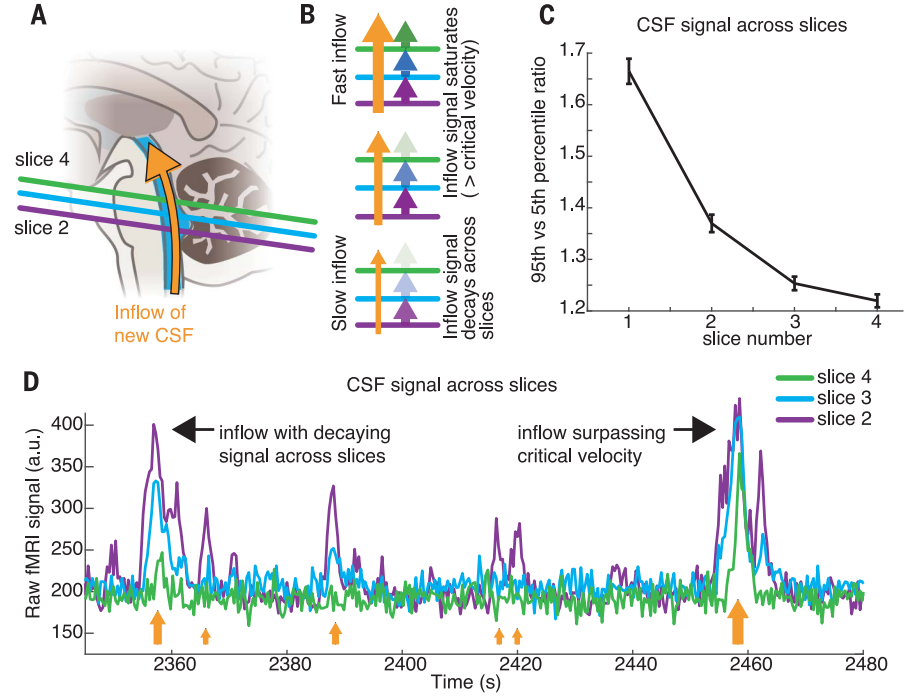


Fig. 2. Ventricle signals correspond to a ~0.05-Hz pulsatile inflow of CSF. (**A**) Schematic of acquisition. New CSF flowing into the imaging volume will generate bright signals. (**B**) Inflow signals will be largest in the bottom slice and decrease in amplitude inwards. If flow exceeds the critical velocity, then CSF in the bottom slice is completely replaced, and signal amplitudes are large in inner slices as well. (**C**) Mean amplitude across slices decays in ascending slices. Error bars are standard error across all sleep segments, with the ROI present in four contiguous slices (n = 129 segments, 11 participants). (**D**) Example time series from the bottom slices of the imaging volume in the fourth ventricle demonstrates the largest signal in the lower slices (e.g., second) and smaller signals in higher slices (e.g., fourth). Orange arrows schematically illustrate flow velocity (larger arrows denote higher velocity), and black arrows point out individual events.

Downloaded from http://science.sciencemag.org/ on November 5, 2019

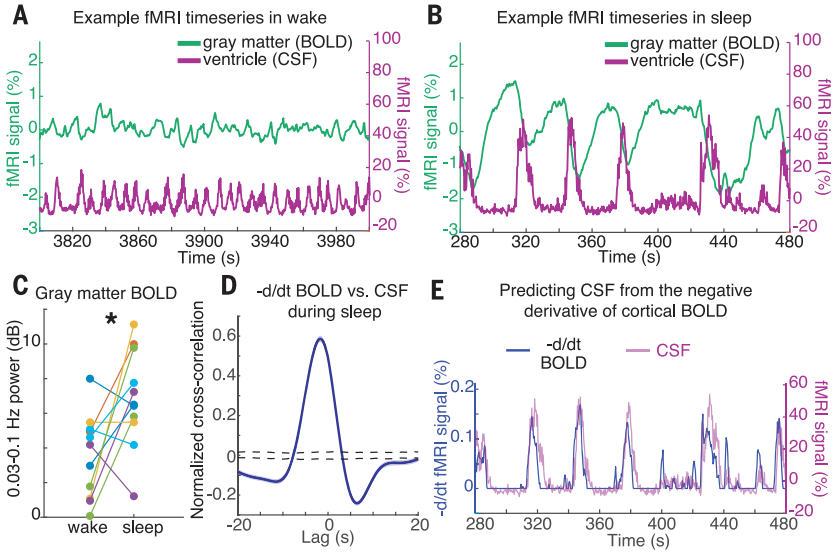


Fig. 3. CSF flow oscillations are anticorrelated to a hemodynamic oscillation in the cortical gray matter that appears during sleep, with CSF flow increasing when blood volume decreases. (**A**) Example time series of the cortical gray-matter BOLD signal and the mean CSF signal from one participant. During wake, signals are low-amplitude and synchronized to respiration (0.25 Hz). (**B**) During sleep, a large-amplitude BOLD oscillation appears, and its time course is coupled to the ventricle CSF signal (~0.05 Hz). (**C**) Mean cortical gray-matter BOLD signal power increases during sleep (n = 11 participants for pairwise test). (**D**) Mean cross-correlation between the zero-thresholded negative derivative of BOLD and CSF signals shows strong correlation (n = 176 segments, 13 participants). The shaded blue region indicates the standard error across segments; the black dashed line denotes the 95% interval of shuffled distribution. (**E**) Example time series showing the correlation, suggesting that CSF flows up the fourth ventricle when cerebral blood volume decreases.

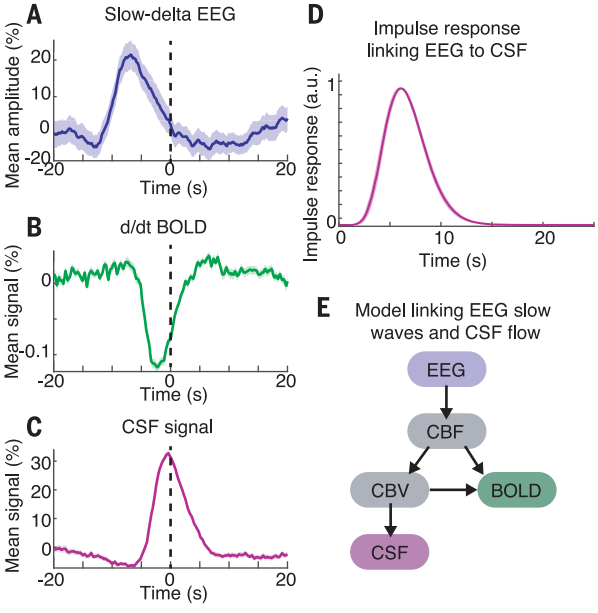


Fig. 4. EEG slow-delta waves are coupled to and precede BOLD and CSF oscillations. (**A**) Mean amplitude envelope of slow-delta EEG, (**B**) mean derivative of BOLD signals, and (**C**) mean CSF signal, all locked to the peaks of CSF waves during sleep. The shaded region represents the standard error across peak-locked trials (n = 123 peaks). (**D**) Calculated impulse response of the CSF signal to the EEG envelope shows a time course similar to that of previously established hemodynamic models. Shading indicates standard deviation across model folds. (**E**) Diagram of model linking the time course of neural activity to CSF flow. Variables include CBF and cerebral blood volume (CBV).

and found that convolving the EEG with this impulse response yielded a significant prediction of CSF dynamics (fig. S5A) (zero-lag r = 0.23, cross-validated r = 0.22 ± 0.07). To examine whether this EEG-CSF coupling was specifically linked to the appearance of CSF waves during sleep, we tested how well the EEG predicted CSF dynamics in the segments with the largest CSF waves. Performance was higher in segments with larger CSF waves (fig. S5, B and C) (zero-lag r = 0.54, P < 0.001, shuffling), suggesting that the EEG was more strongly linked to CSF in the sleep segments richest in CSF waves.

The coherent dynamics of the EEG, BOLD, and CSF signals thus exhibited a specific timing sequence, with neural rhythms preceding subsequent BOLD and CSF waves. To test whether these observed correlations and delay patterns could arise from biophysical mechanisms, we constructed a computational model using established models of hemodynamic coupling (Fig. 4E and fig. S6). We extracted the envelope of slow-delta EEG, expected to correlate with decreases in cerebral blood flow (CBF) signal because of its associated suppression of neural activity (*2*). The neural oscillation was then used to predict the time courses of blood flow, blood volume, and CSF (*33*, *34*). The model first used previously reported physiological parameters (*35*), with no additional parameter fitting, to test whether our observations were consistent with established biophysical coupling between these signals. This model performed as well as the best-fit impulse response [zero-lag R = 0.22; CI across segments = (0.16, 0.27); P < 0.001, shuffling] (fig. S6 and S7). The model prediction was significantly larger than the maximal correlation between the original EEG envelope and CSF across all lags (maximal r = 0.15, P < 0.05), demonstrating that our data were consistent with biophysical coupling between neural activity, hemodynamics, and CSF.

We conclude that human sleep is associated with large coupled low-frequency oscillations in neuronal activity, blood oxygenation, and CSF flow. Although electrophysiological slow waves are known to play important roles in cognition (*1*), our results suggest that they may also be linked to the physiologically restorative effects of sleep, as slow neural activity is followed by brain-wide pulsations in blood volume and CSF flow.

These results address a key missing link in the neurophysiology of sleep. The macroscopic changes in CSF flow that we identified are expected to alter waste clearance, as pulsatile fluid dynamics can increase mixing and diffusion (*20*, *21*, *36*). Neurovascular coupling has been proposed to contribute to clearance (*37*), but why it would cause higher clearance rates during sleep was not known. Our study suggests slow neural and hemodynamic oscillations

Downloaded from http://science.sciencemag.org/ on November 5, 2019

as a possible contributor to this process, in concert with other physiological factors. Studies in animals could next test for causal relationships between these neural and physiological rhythms.

Our identification of sleep-associated CSF fluid dynamics also suggests a potential biomarker to be explored in clinical conditions associated with sleep disturbance. Memory impairment in aging is associated with suppressed slow waves (*38*); our model suggests that this slow-wave loss would, in turn, be associated with decreased CSF flow. Furthermore, our results hint at a potential bridge between recent findings that tau CSF levels and amyloid beta depend on sleep and neural activity (*39*–*41*) and that oscillatory neural activity leads to reduced tau (*42*)—coherent neural activity might signal higher protein aggregate clearance. Taken together, our results identify waves of CSF flow that appear during sleep and show that slow rhythms in neural activity are interlinked with these CSF waves, with hemodynamic oscillations as an intermediate mechanism through which these processes are coupled.

REFERENCES AND NOTES

1. S. Diekelmann, J. Born, *Nat. Rev. Neurosci.* **11**, 114–126 (2010).
2. M. Steriade, A. Nuñez, F. Amzica, *J. Neurosci.* **13**, 3252–3265 (1993).
3. M. Massimini, R. Huber, F. Ferrarelli, S. Hill, G. Tononi, *J. Neurosci.* **24**, 6862–6870 (2004).
4. L. Marshall, H. Helgadóttir, M. Mölle, J. Born, *Nature* **444**, 610–613 (2006).
5. A. Destexhe, D. Contreras, in *Sleep and Anesthesia*, A. Hutt, Ed. (Springer, 2011), p. 258.
6. E. J. W. Van Someren, Y. D. Van Der Werf, P. R. Roelfsema, H. D. Mansvelder, F. H. L. da Silva, *Prog. Brain Res.* **193**, 3–15 (2011).
7. V. V. Vyazovskiy, K. D. Harris, *Nat. Rev. Neurosci.* **14**, 443–451 (2013).
8. B. O. Watson, D. Levenstein, J. P. Greene, J. N. Gelinas, G. Buzsáki, *Neuron* **90**, 839–852 (2016).
9. M. Fukunaga *et al*., *Magn. Reson. Imaging* **24**, 979–992 (2006).
10. S. G. Horovitz *et al*., *Hum. Brain Mapp.* **29**, 671–682 (2008).
11. A. Mitra, A. Z. Snyder, E. Tagliazucchi, H. Laufs, M. E. Raichle, *eLife* **4**, e10781 (2015).
12. M. Boly *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5856–5861 (2012).
13. T. T. Dang-Vu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15160–15165 (2008).
14. C. Kaufmann *et al*., *Brain* **129**, 655–667 (2006).
15. L. J. Larson-Prior *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4489–4494 (2009).
16. L. Xie *et al*., *Science* **342**, 373–377 (2013).
17. L. M. Hablitz *et al*., *Sci. Adv.* **5**, eaav5447 (2019).
18. S. Dreha-Kulaczewski *et al*., *J. Neurosci.* **35**, 2485–2491 (2015).
19. V. Kiviniemi *et al*., *J. Cereb. Blood Flow Metab.* **36**, 1033–1045 (2016).
20. D. Schley, R. Carare-Nnadi, C. P. Please, V. H. Perry, R. O. Weller, *J. Theor. Biol.* **238**, 962–974 (2006).
21. J. J. Iliff *et al*., *J. Neurosci.* **33**, 18190–18199 (2013).
22. H. Mestre *et al*., *Nat. Commun.* **9**, 4878 (2018).
23. I. F. Harrison *et al*., *eLife* **7**, e34028 (2018).
24. P. S. Özbay *et al*., *Neuroimage* **176**, 541–549 (2018).
25. A. Scouten, R. T. Constable, *Magn. Reson. Med.* **59**, 308–315 (2008).
26. S. K. Piechnik, J. Evans, L. H. Bary, R. G. Wise, P. Jezzard, *Magn. Reson. Med.* **61**, 579–586 (2009).
27. S. Ogawa, T. M. Lee, A. R. Kay, D. W. Tank, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9868–9872 (1990).
28. K. K. Kwong *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5675–5679 (1992).
29. N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, A. Oeltermann, *Nature* **412**, 150–157 (2001).
30. B. J. He, M. E. Raichle, *Trends Cogn. Sci.* **13**, 302–309 (2009).
31. C. Mateo, P. M. Knutsen, P. S. Tsai, A. Y. Shih, D. Kleinfeld, *Neuron* **96**, 936–948.e3 (2017).
32. M. L. Schölvinck, A. Maier, F. Q. Ye, J. H. Duyn, D. A. Leopold, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10238–10243 (2010).
33. R. B. Buxton, K. Uludağ, D. J. Dubowitz, T. T. Liu, *Neuroimage* **23** (suppl. 1), S220–S233 (2004).
34. K. J. Friston, A. Mechelli, R. Turner, C. J. Price, *Neuroimage* **12**, 466–477 (2000).
35. A. B. Simon, R. B. Buxton, *Neuroimage* **116**, 158–167 (2015).
36. M. Asgari, D. de Zélicourt, V. Kurtcuoglu, *Sci. Rep.* **6**, 38635 (2016).
37. A. K. Diem, R. O. Carare, R. O. Weller, N. W. Bressloff, *PLOS ONE* **13**, e0205276 (2018).
38. B. A. Mander *et al*., *Nat. Neurosci.* **16**, 357–364 (2013).
39. J. K. Holth *et al*., *Science* **363**, 880–884 (2019).
40. E. Shokri-Kojori *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4483–4488 (2018).
41. J.-E. Kang *et al*., *Science* **326**, 1005–1007 (2009).
42. H. F. Iaccarino *et al*., *Nature* **540**, 230–235 (2016).
43. N. E Fultz *et al*., Sleep-wake imaging ROI time series. Figshare (2019); doi: 10.6084/m9.figshare.9992054.

ACKNOWLEDGMENTS

We thank T. Witzel for contributing to the imaging, J. Voigts and C. Robertson for comments, and S. Williams, K. Gupta, and M. Albrecht for assistance. **Funding:** This work was supported by the National Institutes of Health (R00-MH111748, R01-EB019437, P41-EB015896, R01-MH111419, R01-MH048832, R01-EB024343, R01-MH111438, R21-NS106706, S10-RR023043, and S10-OD010759) and the Martinos Center for Biomedical Imaging. **Author contributions:** All authors contributed to experiment design; L.D.L. and N.E.F. collected and analyzed data; L.D.L. wrote the manuscript; and all authors edited the manuscript. **Competing interests:** L.D.L., B.R.R., and J.R.P. are inventors on a provisional patent coversheet. **Data availability:** The ROI and EEG time series reported here are available at Figshare (*43*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6465/628/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*44*–*57*)

View/request a protocol for this paper from *Bio-protocol*.

15 April 2019; accepted 18 September 2019
10.1126/science.aax5440

Downloaded from http://science.sciencemag.org/ on November 5, 2019

Science

Coupled electrophysiological, hemodynamic, and cerebrospinal fluid oscillations in human sleep

Nina E. Fultz, Giorgio Bonmassar, Kawin Setsompop, Robert A. Stickgold, Bruce R. Rosen, Jonathan R. Polimeni and Laura D. Lewis

Science **366** (6465), 628-631.
DOI: 10.1126/science.aax5440

Fluid dynamics during sleep

During non–rapid eye movement sleep, low-frequency oscillations in neural activity support memory consolidation and neuronal computation. Sleep is also associated with increased interstitial fluid volume and clearance of metabolic waste products. It is unknown why these processes co-occur and how they are related. Fultz *et al.* simultaneously measured electrophysiological, hemodynamic, and flow signals in the human brain (see the Perspective by Grubb and Lauritzen). Large oscillations of fluid inflow to the brain appeared during sleep and were tightly coupled to functional magnetic resonance imaging signals and entrained to electroencephalogram slow waves. Slow oscillatory neuronal activity thus leads to oscillations in blood volume, drawing cerebrospinal fluid into and out of the brain.

Science, this issue p. 628; see also p. 572

ARTICLE TOOLS http://science.sciencemag.org/content/366/6465/628

SUPPLEMENTARY MATERIALS http://science.sciencemag.org/content/suppl/2019/10/30/366.6465.628.DC1

RELATED CONTENT
http://science.sciencemag.org/content/sci/366/6465/572.full
http://stm.sciencemag.org/content/scitransmed/11/474/eaau6550.full
http://stm.sciencemag.org/content/scitransmed/11/514/eaax2014.full
http://stm.sciencemag.org/content/scitransmed/5/183/183ra56.full
http://stm.sciencemag.org/content/scitransmed/5/179/179ra44.full

REFERENCES This article cites 53 articles, 15 of which you can access for free
http://science.sciencemag.org/content/366/6465/628#BIBL

PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

Downloaded from http://science.sciencemag.org/ on November 5, 2019

Behavioral/Systems/Cognitive

Focusing Effect of Acetylcholine on Neuroplasticity in the Human Motor Cortex

Min-Fang Kuo,[1] Jan Grosch,[1] Felipe Fregni,[2] Walter Paulus,[1] and Michael A. Nitsche[1]

[1]Department of Clinical Neurophysiology, Georg-August-University Göttingen, 37075 Göttingen, Germany, and [2]Harvard Center for Noninvasive Brain Stimulation, Boston, Massachusetts 02215

Cholinergic neuromodulation is pivotal for arousal, attention, and cognitive processes. Loss or dysregulation of cholinergic inputs leads to cognitive impairments like those manifested in Alzheimer's disease. Such dysfunction can be at least partially restored by an increase of acetylcholine (ACh). In animal studies, ACh selectively facilitates long-term excitability changes induced by feed-forward afferent input. Consequently, it has been hypothesized that ACh enhances the signal-to-noise ratio of input processing. However, the neurophysiological foundation for its ability to enhance cognition in humans is not well documented. In this study we explore the effects of rivastigmine, a cholinesterase inhibitor, on global and synapse-specific forms of cortical plasticity induced by transcranial direct current stimulation (tDCS) and paired associative stimulation (PAS) on 10–12 healthy subjects, respectively. Rivastigmine essentially blocked the induction of the global excitability enhancement elicited by anodal tDCS and revealed a tendency to first reduce and then stabilize cathodal tDCS-induced inhibitory aftereffects. However, ACh enhanced the synapse-specific excitability enhancement produced by facilitatory PAS and consolidated the inhibitory PAS-induced excitability diminution. These findings are in line with a cholinergic focusing effect that optimizes the detection of relevant signals during information processing in humans.

Key words: **neuroplasticity; acetylcholine; transcranial direct current stimulation; paired associative stimulation; motor cortex; human**

Introduction

Extensive evidence concerning cholinergic modulation of several cognitive functions supports an important role of acetylcholine (ACh) in arousal, attention, learning, and memory formation (Gold, 2003; Sarter et al., 2003). In Alzheimer's disease, enhancing cerebral ACh level has been shown to improve impaired learning and memory functions caused by cholinergic dysfunction. With regard to its specific functional properties, neurophysiological data from animal studies reveal dual neuromodulatory effects of ACh on cortical excitability and synaptic plasticity (Rasmusson, 2000; Gu, 2002). Cholinergic blockade has been shown to reduce long-term potentiation (LTP), whereas cholinergic agonists enhance LTP in the hippocampus, piriform cortex, and neocortex (Blitzer et al., 1990; Brocher et al., 1992; Hasselmo and Barkai, 1995). In humans, use-dependent plasticity of the motor cortex is facilitated by an acetylcholinesterase inhibitor and blocked by a cholinergic antagonist (Sawaki et al., 2002; Meintzschel and Ziemann, 2006). However, it is also reported that ACh enhanced long-term depression (LTD) induced with paired-pulse stimulation in the rat visual cortex (Kirkwood et al., 1999). Furthermore, it suppressed excitatory glutamatergic synaptic transmission via presynaptic inhibition at intrinsic, recurrent synapses but not afferent fiber synapses (Hasselmo and Bower, 1992; Hasselmo et al., 1995; Vogt and Regehr, 2001). This suggests a differential, activity-dependent cholinergic modification of neural networks in which ACh facilitates the detection of incoming afferent inputs, whereas it decreases intrinsic feedback excitability, thereby focusing the encoding of relevant, associated information processing (Blokland et al., 1992; Winters and Bussey, 2005).

To test the focusing action of ACh on neuroplasticity in humans, two protocols of brain stimulation were introduced in the present study. In the paired associative stimulation (PAS) protocol, repetitive peripheral nerve stimulation is paired with transcranial magnetic stimulation (TMS) of the human motor cortex (Stefan et al., 2000). It is postulated that PAS-induced excitability changes share the features of associative synaptic LTP and LTD, depending on the sequence of the near-synchronous pair of stimuli from different stimulation modalities in the motor cortex (Stefan et al., 2000; Wolters et al., 2003), which parallels the spike-timing-dependent rule for Hebbian LTP and LTD induction in animal studies (Dan and Poo, 2004). Facilitatory PAS (PAS25) with peripheral nerve stimulation applied 25 ms earlier than TMS pulse in M1 results in synchronous activation of motor cortical neurons by the afferent somatosensory stimulus and motor cortex TMS and thus enhances cortical excitability. However, inhibitory PAS (PAS10) diminishes cortical excitability with the interstimuli interval 10 ms, with which the somatosensory input reaches the motor cortex relevantly later than the TMS pulse, thereby inducing asynchronous stimulation on motor cortical neurons. PAS should specifically induce neuroplasticity in somatosensory-motor cortical synapses. In contrast, transcranial

Received May 17, 2007; revised Oct. 18, 2007; accepted Nov. 1, 2007.

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Program "Nicotine (SPP 1226)." M.-F.K. was supported by European Graduiertenkolleg 632, "Neuroplasticity: From Molecules to Systems," which was funded by DFG.

Correspondence should be addressed to Michael A. Nitsche, Department of Clinical Neurophysiology, Georg-August-University Göttingen, Robert-Koch-Strasse 40, 37075 Göttingen, Germany. E-mail: mnitsch1@gwdg.de.

DOI:10.1523/JNEUROSCI.4104-07.2007

Copyright © 2007 Society for Neuroscience 0270-6474/07/2714442-06$15.00/0

direct current stimulation (tDCS) encompasses the global modulation of cortical network plasticity by application of weak direct currents through the surface of the scalp. Anodal tDCS enhances cortical excitability, whereas cathodal tDCS diminishes it for up to 1 h after the end of stimulation (Nitsche and Paulus, 2000, 2001; Nitsche et al., 2003a). The primary mechanism is a modulation of the resting membrane potential, and the resulting polarity-specific excitability changes subsequently induce changes of synaptic strength, which are, however, not restricted to specific synaptic connections (Bindman et al., 1964; Purpura and McMurtry, 1965). Both plasticity-inducing protocols accomplish long-lasting, NMDA receptor-dependent excitability changes (Liebetanz et al., 2002; Stefan et al., 2002; Nitsche et al., 2003b). The main difference lies in the synapse-specific focal effects of PAS: whereas the plasticity induced by tDCS is relatively nonfocal and not synapse-specific because it is thought to change cortical excitability under the whole area covered by the relatively large stimulation electrode, the plasticity induced by PAS is restricted to the intercortical connections between the somatosensory and motor cortex. According to the focusing hypothesis of ACh, it should selectively enhance and consolidate specific synaptic modifications induced by PAS, while depressing global ones accomplished by tDCS, to sharpen the signal-to-noise ratio in human cortical networks.

Materials and Methods

Subjects. Ten to twelve healthy subjects [tDCS experiment: six men, six women, aged (mean ± SD) 24 ± 3 years; PAS25 experiment: four women and six men, aged 28 ± 4 years; PAS10 experiment: 5 women and 5 men, aged 27 ± 4 years], without receiving acute or chronic medication, participated in each experiment. The groups did not differ significantly with regard to age and gender. Both studies were approved by the ethics committee of the University of Goettingen, and we conform to the Declaration of Helsinki. All subjects had given written informed consent. Because we were interested primarily in the physiological effects of ACh in this study, we recruited relatively young subjects to guarantee the comparability of the results with former pharmacological studies on neuroplasticity and to avoid including subjects with subclinical (e.g., microvascular) brain lesions, which might have influenced our results unintentionally.

Transcranial direct current stimulation. tDCS was performed with a pair of saline-soaked surface sponge electrodes (35 cm^2) with one of the electrodes placed over the representational area of the right abductor digiti minimi muscle (ADM), as determined by TMS, and the other electrode above the right orbit as reference. The currents ran continuously for 13 min (anodal tDCS) or 9 min (cathodal tDCS) with an intensity of 1 mA. In previous studies, these stimulation durations have been shown to induce aftereffects of tDCS lasting for ~1 h (Nitsche and Paulus, 2001; Nitsche et al., 2003a).

Paired associative stimulation. Peripheral nerve stimulation was applied on the right ulnar nerve at the level of the wrist with a Digitimer D185 stimulator (Digitimer, Welwyn Garden City, UK). Single-pulse TMS was delivered over the representing area of the right ADM. Each TMS pulse, at an intensity eliciting a muscle evoked potential of ~1 mV peak-to-peak amplitude, was preceded by an ulnar nerve stimulus with an interval of 25 ms (PAS25) or 10 ms (PAS10) using a standard stimulation block (cathodal proximal) at a stimulation width of 200 μs and stimulation intensity of 300% of the perceptual threshold, defined as the lowest intensity of the stimuli that is perceivable by the subject. Ninety pairs were applied at 0.05 Hz over 30 min, which has been shown to induce long-lasting excitability changes in the motor cortex (Stefan et al., 2000).

Pharmacological interventions. Rivastigmine (3 mg) or equivalent placebo drugs were taken by the subjects 2 h before the start of the intervention (Kennedy et al., 1999). This dose was chosen to minimize drug-induced side effects, but to enhance the cholinergic level of the CNS effectively. The experimental sessions were performed in a randomized order and were separated by at least 1 week to avoid cumulative drug or stimulation effects.

Measurement of motor cortical excitability. TMS-elicited muscle-evoked potentials (MEPs) were recorded to measure excitability changes of the representational motor cortical area of the right ADM. Single-pulse TMS was conducted by a Magstim 200 magnetic stimulator (Magstim Company, Whiteland, Dyfed, UK) with a figure-eight-shaped magnetic coil (diameter of one winding, 70 mm; peak magnetic field, 2.2 tesla). The coil was held tangentially to the skull, with the handle pointing backward and laterally at an angle of 45° from midline. The optimal position was defined as the site at which stimulation resulted consistently in the largest MEPs. Surface EMG was recorded from the right ADM with Ag-AgCl electrodes in a belly-tendon montage. The signals were amplified and filtered with a time constant of 10 ms and a low-pass filter of 2.5 kHz and then digitized at an analog-to-digital rate of 5 kHz and further relayed into a laboratory computer using the Signal software and CED 1401 hardware (Cambridge Electronic Design, Cambridge, UK). The intensity was adjusted to elicit baseline MEPs of, on average, 1 mV peak-to-peak amplitude and was kept constant for the poststimulation assessment unless adjusted (see below).

Experimental procedures. The experiments were conducted in a repeated measurement design. For the tDCS experiment, a complete crossover design was chosen. For the PAS experiments, separate subject groups participated in the PAS10 and PAS25 experiments. Subjects were seated comfortably in a reclining chair. First the optimal position of the magnetic coil for eliciting MEPs in the resting ADM was assessed over the left motor cortex, and 20 MEPs were recorded for the first baseline. Two hours after intake of the medication, a second baseline was determined to control for a possible influence of the drug on cortical excitability and adjusted if necessary. Subjects were not aware about and could not distinguish between the specific stimulation protocols used, nor were they informed about the medication administered in a specific session.

In experiment 1 with tDCS, one of the DC electrodes, to which in the following the terms cathodal or anodal tDCS refer, was fixed at the cortical representational area of ADM as defined during the first baseline recording, and the other one was fixed at the contralateral forehead area above the right orbit. Direct currents were applied on 12 subjects for 9 min (cathodal) or 13 min (anodal). After cessation of tDCS, 20 MEPs were recorded at 0.25 Hz every 5 min for half an hour and then every 30 min until 2 h after the end of DC stimulation, because tDCS-induced aftereffects without medication will not last longer than this period of time (Nitsche and Paulus, 2000; Nitsche et al., 2003a). Because of the relatively diurnal stability of corticospinal excitability, we did not expect additional excitability changes after MEPs returned to baseline level again. Only for the rivastigmine conditions, TMS recordings were performed at four additional time points: same day evening, next morning, next noon, and next evening.

In experiment 2 with PAS, the interventional PAS protocol as described above was used on 10 subjects. TMS recording procedures were the same as described above.

Data analysis and statistics. MEP amplitude means were calculated first individually and then interindividually for each time bin including both baseline values. The postintervention MEPs were normalized and are given as ratios of the baseline determined immediately before intervention (tDCS/PAS).

In the tDCS experiment, a repeated-measure ANOVA for the time bins up to 120 min after tDCS was calculated with the within-subject factors time course, current stimulation (anodal and cathodal tDCS), drug condition (rivastigmine vs placebo), and the dependent variable MEP amplitude. For PAS experiment, we performed a repeated-measure ANOVA with PAS (PAS25 and PAS10) as between-subject factor and within-subject factors drug condition (rivastigmine vs placebo) and time course for MEPs up to 120 min after intervention. If appropriate, *post hoc* Student's *t* tests (paired samples, two-tailed, $p < 0.05$, not adjusted) were performed to determine whether the MEP amplitudes before and after the interventional brain stimulations differed in each intervention condition and whether those differences depended on the drug conditions.

Table 1. Results of the ANOVAs

	Parameters	df	*F* value	*p* value
Experiment 1 (tDCS)	TDCS	1	40.585	< 0.001*
	Drug	1	3.650	0.082
	Time course	10	1.700	0.099
	tDCS × drug	1	3.417	0.092
	tDCS × time course	10	1.868	0.066
	Drug × time course	10	0.290	0.976
	tDCS × drug × time course	10	4.206	< 0.001*
Experiment 2 (PAS)	PAS	1	73.006	< 0.001*
	Drug	1	0.140	0.712
	Time course	10	1.556	0.123
	Drug × PAS	1	23.886	< 0.001*
	Time course × PAS	10	9.934	< 0.001*
	Drug × time course	10	2.207	0.019*
	Drug × time course × PAS	10	3.208	0.001*

In both experiments, the ANOVAs encompass the time course up to 120 min after tDCS or PAS, because the remaining time points were only measured for the rivastigmine conditions. "Drug" represents rivastigmine and placebo, whereas "tDCS" indicates anodal and cathodal polarity, and "PAS" represents PAS25 and PAS10. $p \leq 0.05$.

Additional *post hoc* tests (Student's *t* tests, $p < 0.05$) were performed to explore whether rivastigmine modified baseline MEPs.

Results

Three subjects experienced prominent side effects under rivastigmine (nausea, vomiting, and dizziness) and were excluded from the experiment. Two of the participating subjects complained about slight nausea after taking rivastigmine, but the symptoms subsided before the neuroplasticity-inducing intervention. Baseline MEP amplitudes did not differ significantly before and after drug intake in all conditions. Absolute baseline MEP amplitudes were not different in all medication and stimulation subgroups (Student's *t* tests, paired, two-tailed, $p > 0.05$).

Effects of rivastigmine on tDCS-induced motor cortex excitability shifts (experiment 1)

The ANOVA revealed a significant main effect of tDCS ($F = 40.585$; $p < 0.001$) and tDCS × drug × time course ($F = 4.206$; $p < 0.001$) (Table 1). In the PLC conditions, the anodal tDCS-induced excitability increase stayed significant until 30 min after tDCS, and the cathodal tDCS-induced inhibition lasted until 60 min after DC stimulation. As revealed by the *post hoc t* tests (paired, two-tailed, $p < 0.05$), rivastigmine initially abolished the induction of both the anodal tDCS-elicited excitability enhancement and the cathodal tDCS-elicited excitability diminution. However, a delayed, consolidated inhibition induced by cathodal tDCS was observed in the rivastigmine condition, compared with placebo conditions. The decrease of excitability generated by cathodal tDCS under rivastigmine remained significant for 2 h after tDCS, whereas it returned to baseline level after 60 min in the PLC condition (Fig. 1).

Effects of rivastigmine on PAS-induced motor cortex excitability shifts (experiment 2)

The ANOVA displayed a significant main effect of PAS ($F = 73.006$; $p < 0.001$). The interactions between drug × PAS ($F = 23.886$; $p < 0.001$), time course × PAS ($F = 9.934$; $p < 0.001$), drug × time course ($F = 2.207$; p 0.019), and drug × time course × PAS ($F = 3.208$; $p < 0.001$) were also significant (Table 1). In the PAS25 experiment, the excitatory shift of MEP amplitudes returned to baseline 25 min after PAS in the placebo medication condition, as revealed by Student's *t* tests (paired, two-tailed, $p < 0.05$), whereas rivastigmine enhanced and prolonged the excitatory effects of PAS25 (Fig. 2). In the PAS10 experiment, the inhibitory effect without medication lasted for half an hour. Rivastigmine further enhanced and prolonged the inhibition until the same day evening after intervention (Fig. 3).

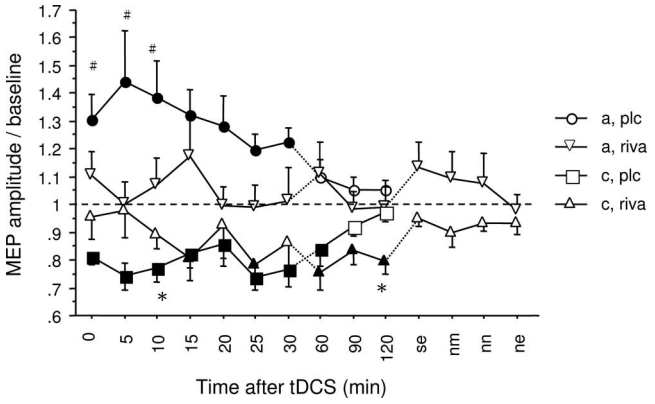


Figure 1. Cholinergic modulation of global cortical plasticity induced by tDCS. Rivastigmine abolished the induction of anodal tDCS-elicited excitability increases, as recorded by TMS-evoked MEP amplitudes. Additionally, rivastigmine initially diminished the excitability reduction induced by cathodal tDCS under rivastigmine. However, the respective excitability diminution was later consolidated. Filled symbols indicate significant deviations from baseline with regard to each drug condition. Hash symbols indicate significant differences in anodal tDCS-induced excitability changes between placebo and rivastigmine conditions; asterisks represent significant differences in inhibition caused by cathodal stimulation between the placebo and rivastigmine medication conditions (Student's *t* test, two-tailed, repeated measures; $^{\#}p < 0.05$). a, Anodal; c, cathodal; plc, placebo; riva, rivastigmine; se, same evening; nm, next morning; nn, next noon; ne, next evening. Error bars indicate SEM.

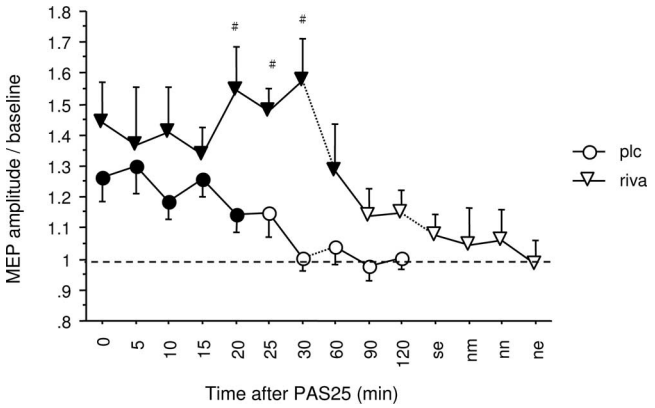


Figure 2. The PAS25-induced synapse-specific excitability enhancement is facilitated by ACh. The PAS-induced excitability enhancement was increased and prolonged under rivastigmine until 30 min after PAS, whereas MEP amplitudes returned to baseline 25 min after PAS in the placebo condition. Hash symbols represent significant differences between placebo and rivastigmine conditions; filled symbols indicate significant deviations from baseline with regard to each drug condition (Student's *t* test, two-tailed, repeated measures; $^{\#}p < 0.05$). plc, Placebo; riva, rivastigmine; se, same evening; nm, next morning; nn, next noon; ne, next evening. Error bars indicate SEM.

Discussion

The present study demonstrates that ACh enhances the synapse-specific cortical excitability increase induced by PAS25 and consolidates the PAS10-induced reduction of motor cortical excitability, whereas it prevents global excitatory aftereffects produced by anodal tDCS. ACh also delayed the induction of the cathodal tDCS-elicited excitability decrease and slightly prolonged its overall duration. Because MEP amplitudes were not modified by rivastigmine in the dosage applied alone, we have no evidence for a direct cholinergic influence of the drug on corticospinal excit-

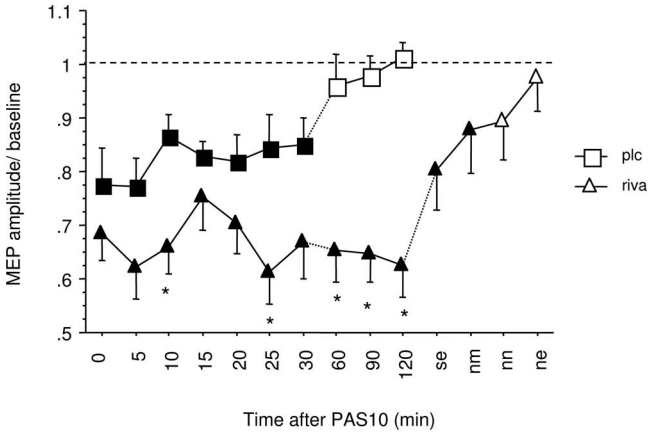


Figure 3. The PAS10-induced synapse-specific excitability diminution is consolidated by ACh. In the placebo condition, the inhibitory effect of PAS10 lasted for approximately half an hour, whereas rivastigmine enhanced the inhibition and further prolonged it until same day evening of the intervention. Significant differences between placebo and rivastigmine are shown with asterisks. Filled symbols indicate significant deviations from baseline with regard to each drug condition (Student's *t* test, two-tailed, repeated measures, $p < 0.05$). plc, Placebo; riva, rivastigmine; se, same evening; nm, next morning; nn, next noon; ne, next evening. Error bars indicate SEM.

ability. The results support the hypothesis of a focusing effect of ACh on neuroplasticity of cortical networks. ACh not only increased selectively the efficacy of synapse-specific excitability-enhancing neuroplasticity, but also prolonged the excitability diminution in case of asynchronous synapse-specific inputs and suppressed global excitability enhancements. By these mechanisms, ACh is well suited to improve the signal-to-noise ratio and to refine information processing in neural networks.

ACh diminishes tDCS-driven neuroplasticity

At first glance, the inhibitory effect of rivastigmine on facilitatory neuroplasticity induced by anodal tDCS seems contradictory to the results obtained from animal studies in which LTP was facilitated by cholinergic stimulation (Brocher et al., 1992; Abe et al., 1994; Hasselmo and Barkai, 1995; Patil et al., 1998). The major conceptual difference between these studies and our tDCS experiment is the manipulation applied for neuroplasticity induction. In slice preparation, the plastic changes induced with either paired or high-frequency suprathreshold electrical stimulation are caused by synchronous activation of synaptic connections. In contrast, tDCS-elicited neuroplasticity is a consequence of increasing general, asynchronous network activity, possibly added by modification of the postsynaptic resting membrane potential (Bindman et al., 1964; Purpura and McMurtry, 1965). The diminution of the anodal tDCS-induced excitability enhancement might thus be because of an ACh-induced decrease of general excitation within global neuronal networks, probably modulated via the cholinergic presynaptic inhibition of excitatory feedback potentials or excitatory transmission at recurrent connections. Moreover, evidence reveals that the suppression of synaptic transmission is selective for recently modified synapses (Linster et al., 2003) but does not apply to silent synapses (Fernandez de Sevilla et al., 2002). Thus it is probable that synapses that are globally modified by tDCS in the present study are more susceptible to cholinergic suppression of synaptic transmission during plasticity induction. A similar effect is also demonstrated by a recent study using a dynamic clamp system to mimic *in-vivo*-like background activities in motor cortical slices, in which cholinergic facilitation of LTP was attenuated in the presence of random background noise (Desai and Walcott, 2006).

However, ACh revealed a tendency to consolidate excitability-diminishing aftereffects generated by cathodal tDCS, although it abolished its initial induction phase. The biphasic effect of rivastigmine on cathodal tDCS-induced plasticity can also be explained by a cholinergic regulation of different neurons within different temporary profiles. The initial blockade could be because of the fast negative modulation of inhibitory neurons (e.g., inhibitory interneurons) known to be induced by ACh (Ji and Dani, 2000), whereas the late-onset and prolonged excitability diminution can be explained by the inhibitory modification of excitatory neurons with the combination of ACh and cathodal tDCS. Further studies are required to test this hypothesis.

Cholinergic consolidation of PAS-induced cortical plasticity

The PAS experiment demonstrates a positive cholinergic modulation of PAS-elicited synaptic-specific plasticity. ACh seems to enhance specifically the ability of synchronous motor cortical input to enhance excitability, whereas the excitability diminution accomplished by asynchronous input is prolonged. It is suggested that PAS25/PAS10 relates to associative LTP/LTD in the human motor cortex (Stefan et al., 2000, 2002; Wolters et al., 2003; Stefan et al., 2006). Because ACh has been shown to facilitate cortical sensory plasticity by enhancing sensory input processing (Rasmusson and Dykes, 1988; Tremblay et al., 1990; Patil et al., 1998) and suppressing irrelevant input (Hasselmo and Barkai, 1995), one might further speculate that rivastigmine specifically improved the efficacy of PAS by (1) enhancing the signal-to-noise ratio, thereby facilitating meaningful information processing as represented by synchronous input; and (2) suppressing non-meaningful input as represented by asynchronous stimulation within neural networks.

Thus, the results of our study are in accordance with the respective animal experiments (Blitzer et al., 1990; Hasselmo and Barkai, 1995; Kirkwood et al., 1999). Moreover, they are concordant with the results of recent behavioral studies in the human motor cortex exploring the effect of ACh modulation on use-dependent plasticity, which was blocked by the ACh antagonist scopolamine (Sawaki et al., 2002) and enhanced by an acetylcholinesterase inhibitor (Meintzschel and Ziemann, 2006). Therefore, our results offer a neurophysiological mechanism of how ACh might improve behavioral plasticity.

It is also notable that the duration of cholinergic effects on cortical plasticity induced by PAS25 and PAS10 is asymmetric. ACh was more effective in stabilizing the PAS10-generated excitability reduction than the PAS25-induced excitability enhancement. This might be because of the generally less efficient PAS25 protocol, as revealed in the placebo condition in the present study, which might have limited the ability of ACh to stabilize neuroplasticity. Alternatively, it was suggested that the direction of cholinergic modulation on synaptic plasticity could be determined by ACh concentration and subtype receptors activation (Kuczewski et al., 2005), which could also explain an asymmetry of ACh effects on different PAS protocols.

Summary of cholinergic modulation in human cortical plasticity

The results of the present study suggest that ACh has fairly specific effects on cortical plasticity. Rivastigmine selectively enhanced the efficacy of PAS25, a synchronous associative stimulation protocol, to increase excitability, whereas it shifted the effects of the remaining plasticity-inducing protocols in an inhib-

itory direction. This indicates a general inhibitory effect of rivastigmine on network excitability, with the exception of PAS25-induced plasticity, and further explains how rivastigmine works as a cognitive enhancer via increasing the signal-to-noise ratio of cortical activity. Associative plasticity is suggested as a neurophysiological correlate of learning and memory formation, and, indeed, excitability-enhancing PAS has been shown to be tightly connected to learning processes (Ziemann et al., 2004; Stefan et al., 2006), thus strengthening the efficacy of synchronous stimuli to enhance excitability might improve learning. However, rivastigmine can reduce "noisy" synaptic modification, as demonstrated in our study via abolishing the excitability-enhancing properties of anodal tDCS, which increases general and most probably not synchronous cortical activity (Bindman et al., 1964; Purpura and McMurtry, 1965), and via strengthening the inhibitory effects of the asynchronous PAS10 protocol. Moreover, rivastigmine enhanced the efficacy of cathodal tDCS to diminish cortical excitability. Cathodal tDCS applied synchronously with PAS25 has been shown to enhance the efficacy of PAS25 to increase cortical excitability (Nitsche et al., 2007), which can be explained by a noise-reducing function of cathodal tDCS-induced inhibition. Thus the enhancing effect of rivastigmine on cathodal tDCS-driven inhibition should also increase the efficacy of meaningful information processing via increasing the signal-to-noise ratio.

A recently conducted study revealed similar effects of L-dopa (L-3,4-dihydroxyphenylalanine) on neuroplasticity (Kuo et al., 2007). Here, dopamine reversed the anodal tDCS-induced excitability enhancement into inhibition and prolonged the excitability diminution caused by cathodal tDCS. However, dopamine enhanced the PAS25-induced synapse-specific increase of cortical excitability. The similarity of the dopamine and ACh effects could be explained by the fact that dopaminergic and cholinergic neurons serve similar functions and are tightly interconnected in neuronal networks (Sarter et al., 1999; Zahm, 2006). They also provide, at least, a theoretical option for the application of dopamine in neurological disorders associated with cognitive deficits. Indeed, dopaminergic medication has been shown to improve learning in healthy subjects and patients after stroke (Floel et al., 2005, 2007).

Together, the results of the present study are in line with cortical cholinergic functions enhancing the contrast of relevant stimuli against background noise or distractors, thereby improving signal processing during information encoding. Therefore, it might strengthen the rational basis for application of cholinesterase inhibitors to improve cognitive functions in patients with Alzheimer's disease or vascular dementia. However, it has to be kept in mind that the results were obtained with healthy young subjects. Further studies are needed to elucidate whether the effect of ACh on neuroplasticity is identical in elderly healthy and demented patients and whether the impact of ACh on neuroplasticity is correlated with clinical outcome.

References

Abe K, Nakata A, Mizutani A, Saito H (1994) Facilitatory but nonessential role of the muscarinic cholinergic system in the generation of long-term potentiation of population spikes in the dentate gyrus in vivo. Neuropharmacology 33:847–852.

Bindman LJ, Lippold OCJ, Redfearn JWT (1964) The action of brief polarizing currents on the cerebral cortex of the rat (1) during current flow and (2) in the production of long-lasting after-effects. J Physiol (Lond) 172:369–382.

Blitzer RD, Gil O, Landau EM (1990) Cholinergic stimulation enhances long-term potentiation in the CA1 region of rat hippocampus. Neurosci Lett 119:207–210.

Blokland A, Honig W, Raaijmakers WG (1992) Effects of intrahippocampal scopolamine injections in a repeated spatial acquisition task in the rat. Psychopharmacology (Berl) 109:373–376.

Brocher S, Artola A, Singer W (1992) Agonists of cholinergic and noradrenergic receptors facilitate synergistically the induction of long-term potentiation in slices of rat visual cortex. Brain Res 573:27–36.

Dan Y, Poo MM (2004) Spike timing-dependent plasticity of neural circuits. Neuron 44:23–30.

Desai NS, Walcott EC (2006) Synaptic bombardment modulates muscarinic effects in forelimb motor cortex. J Neurosci 26:2215–2226.

Fernandez de Sevilla D, Cabezas C, de Prada AN, Sanchez-Jimenez A, Buno W (2002) Selective muscarinic regulation of functional glutamatergic Schaffer collateral synapses in rat CA1 pyramidal neurons. J Physiol (Lond) 545:51–63.

Floel A, Hummel F, Breitenstein C, Knecht S, Cohen LG (2005) Dopaminergic effects on encoding of a motor memory in chronic stroke. Neurology 65:472–474.

Floel A, Garraux G, Xu B, Breitenstein C, Knecht S, Herscovitch P, Cohen LG (2007) Levodopa increases memory encoding and dopamine release in the striatum in the elderly. Neurobiol Aging, in press.

Gold PE (2003) Acetylcholine modulation of neural systems involved in learning and memory. Neurobiol Learn Mem 80:194–210.

Gu Q (2002) Neuromodulatory transmitter systems in the cortex and their role in cortical plasticity. Neuroscience 111:815–835.

Hasselmo ME, Barkai E (1995) Cholinergic modulation of activity-dependent synaptic plasticity in the piriform cortex and associative memory function in a network biophysical simulation. J Neurosci 15:6592–6604.

Hasselmo ME, Bower JM (1992) Cholinergic suppression specific to intrinsic not afferent fiber synapses in rat piriform (olfactory) cortex. J Neurophysiol 67:1222–1229.

Hasselmo ME, Schnell E, Barkai E (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. J Neurosci 15:5249–5262.

Ji D, Dani JA (2000) Inhibition and disinhibition of pyramidal neurons by activation of nicotinic receptors on hippocampal interneurons. J Neurophysiol 83:2682–2690.

Kennedy JS, Polinsky RJ, Johnson B, Loosen P, Enz A, Laplanche R, Schmidt D, Mancione LC, Parris WC, Ebert MH (1999) Preferential cerebrospinal fluid acetylcholinesterase inhibition by rivastigmine in humans. J Clin Psychopharmacol 19:513–521.

Kirkwood A, Rozas C, Kirkwood J, Perez F, Bear MF (1999) Modulation of long-term synaptic depression in visual cortex by acetylcholine and norepinephrine. J Neurosci 19:1599–1609.

Kuczewski N, Aztiria E, Gautam D, Wess J, Domenici L (2005) Acetylcholine modulates cortical synaptic transmission via different muscarinic receptors, as studied with receptor knockout mice. J Physiol (Lond) 566:907–919.

Kuo MF, Paulus W, Nitsche MA (2007) Boosting focally-induced brain plasticity by dopamine. Cereb Cortex, in press.

Liebetanz D, Nitsche MA, Tergau F, Paulus W (2002) Pharmacological approach to the mechanisms of transcranial DC-stimulation-induced after-effects of human motor cortex excitability. Brain 125:2238–2247.

Linster C, Maloney M, Patil M, Hasselmo ME (2003) Enhanced cholinergic suppression of previously strengthened synapses enables the formation of self-organized representations in olfactory cortex. Neurobiol Learn Mem 80:302–314.

Meintzschel F, Ziemann U (2006) Modification of practice-dependent plasticity in human motor cortex by neuromodulators. Cereb Cortex 16:1106–1115.

Nitsche MA, Paulus W (2000) Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. J Physiol 527 Pt 3:633–639.

Nitsche MA, Paulus W (2001) Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. Neurology 57:1899–1901.

Nitsche MA, Nitsche MS, Klein CC, Tergau F, Rothwell JC, Paulus W (2003a) Level of action of cathodal DC polarisation induced inhibition of the human motor cortex. Clin Neurophysiol 114:600–604.

Nitsche MA, Fricke K, Henschke U, Schlitterlau A, Liebetanz D, Lang N,

Henning S, Tergau F, Paulus W (2003b) Pharmacological modulation of cortical excitability shifts induced by transcranial direct current stimulation in humans. J Physiol (Lond) 553:293–301.

Nitsche MA, Roth A, Kuo MF, Fischer AK, Liebetanz D, Lang N, Tergau F, Paulus W (2007) Timing-dependent modulation of associative plasticity by general network excitability in the human motor cortex. J Neurosci 27:3807–3812.

Patil MM, Linster C, Lubenov E, Hasselmo ME (1998) Cholinergic agonist carbachol enables associative long-term potentiation in piriform cortex slices. J Neurophysiol 80:2467–2474.

Purpura DP, McMurtry JG (1965) Intracellular activities and evoked potential changes during polarization of motor cortex. J Neurophysiol 28:166–185.

Rasmusson DD (2000) The role of acetylcholine in cortical synaptic plasticity. Behav Brain Res 115:205–218.

Rasmusson DD, Dykes RW (1988) Long-term enhancement of evoked potentials in cat somatosensory cortex produced by co-activation of the basal forebrain and cutaneous receptors. Exp Brain Res 70:276–286.

Sarter M, Bruno JP, Turchi J (1999) Basal forebrain afferent projections modulating cortical acetylcholine, attention, and implications for neuropsychiatric disorders. Ann NY Acad Sci 877:368–382.

Sarter M, Bruno JP, Givens B (2003) Attentional functions of cortical cholinergic inputs: what does it mean for learning and memory? Neurobiol Learn Mem 80:245–256.

Sawaki L, Boroojerdi B, Kaelin-Lang A, Burstein AH, Butefisch CM, Kopylev L, Davis B, Cohen LG (2002) Cholinergic influences on use-dependent plasticity. J Neurophysiol 87:166–171.

Stefan K, Kunesch E, Cohen LG, Benecke R, Classen J (2000) Induction of plasticity in the human motor cortex by paired associative stimulation. Brain 123 Pt 3:572–584.

Stefan K, Kunesch E, Benecke R, Cohen LG, Classen J (2002) Mechanisms of enhancement of human motor cortex excitability induced by interventional paired associative stimulation. J Physiol (Lond) 543:699–708.

Stefan K, Wycislo M, Gentner R, Schramm A, Naumann M, Reiners K, Classen J (2006) Temporary occlusion of associative motor cortical plasticity by prior dynamic motor training. Cerebral Cortex 16:376–385.

Tremblay N, Warren RA, Dykes RW (1990) Electrophysiological studies of acetylcholine and the role of the basal forebrain in the somatosensory cortex of the cat. II. Cortical neurons excited by somatic stimuli. J Neurophysiol 64:1212–1222.

Vogt KE, Regehr WG (2001) Cholinergic modulation of excitatory synaptic transmission in the CA3 area of the hippocampus. J Neurosci 21:75–83.

Winters BD, Bussey TJ (2005) Removal of cholinergic input to perirhinal cortex disrupts object recognition but not spatial working memory in the rat. Eur J Neurosci 21:2263–2270.

Wolters A, Sandbrink F, Schlottmann A, Kunesch E, Stefan K, Cohen LG, Benecke R, Classen J (2003) A temporally asymmetric Hebbian rule governing plasticity in the human motor cortex. J Neurophysiol 89:2339–2345.

Zahm DS (2006) The evolving theory of basal forebrain functional-anatomical "macrosystems." Neurosci Biobehav Rev 30:148–172.

Ziemann U, Iliac TV, Pauli C, Meintzschel F, Ruge D (2004) Learning modifies subsequent induction of long-term potentiation-like and long-term depression-like plasticity in human motor cortex. J Neurosci 24:1666–1672.

PET Imaging of Dopamine D2 Receptors in Monkey Models of Cocaine Abuse: Genetic Predisposition Versus Environmental Modulation

Michael A. Nader, Ph.D.

Paul W. Czoty, Ph.D.

Objective: Animals self-administer many of the drugs that humans abuse, including cocaine. This article describes studies using preclinical animal models to differentiate the influences of neurobiological predisposition from environmental modulation of cocaine addiction, including studies from the authors' laboratory using nonhuman primates.

Method: Addiction is described in terms of vulnerability, maintenance, and abstinence. This review focuses on dopamine receptor function, in particular that of the D2-like receptors, as measured by the noninvasive imaging procedure positron emission tomography. Findings from human studies of addiction and animal models are reviewed.

Results: There appears to be an inverse relationship between D2 receptor availability and vulnerability to the reinforcing effects of cocaine. Environmental variables can increase or decrease D2 receptor binding in an orderly fashion, and the resulting changes in D2 function influence the vulnerability to abuse cocaine. In maintenance, chronic cocaine exposure produces decreases in D2 receptor binding, which may be a mechanism that contributes to continued drug use. Finally, during abstinence there are individual differences in rates of recovery of D2 receptor availability.

Conclusions: The goal of the preclinical research described in this review is to achieve a better understanding of individual differences in susceptibility and vulnerability to the reinforcing effects of cocaine. It is clear that the development of novel animal models will extend our understanding of the neurobiological basis of drug addiction to include a greater appreciation of the role of environmental factors in affecting predisposition, mediating continued drug use, and triggering relapse.

(Am J Psychiatry 2005; 162:1473–1482)

Recent estimates indicate that the number of new cocaine users per year more than doubled from 1992 (0.5 million) to 2002 (1.1 million) (1). From 1994 to 2001, visits to hospital emergency rooms that involved cocaine increased by 22% (2). Despite more than three decades of research into the molecular, cellular, and behavioral effects of cocaine, no pharmacotherapy for cocaine abuse has demonstrated sufficient safety and effectiveness for widespread clinical use (3, 4). As astutely pointed out by Leshner (5), most people see drug abuse and addiction as a social problem, and as such, the problem is frequently handled only with social solutions, in particular through the criminal justice system. Recent reports indicate that the 40%–60% of people seeking treatment are ordered to do so by the courts or corrections system (see reference 6), suggesting that treatment does not occur until drug use has led to substantial adverse consequences to the individual. However, if drug addiction is viewed as a brain disease, then addiction should be amenable to treatments, just as depression, schizophrenia, and anxiety are treatable. As Leshner stated,

> Understanding addiction as a brain disease explains in part why historic policy strategies focusing solely on the social or criminal justice aspects of drug use and addiction have been unsuccessful. They are missing at least half of the issue. If the brain is the core of the problem, attending to the brain needs to be a core part of the solution. (5, p. 47)

In response to growing evidence, the World Health Organization recently issued a report focusing on the role of the brain in mediating drug dependence (7). The goal of the present review is to highlight some of the brain imaging studies of cocaine abuse with a special emphasis on the use of animal models to help us understand the human condition.

Use of Brain Imaging Techniques in Cocaine Abuse Research

Cocaine is an indirectly acting monoamine agonist, binding to the dopamine, serotonin, and norepinephrine transporters (8, 9). The majority of research on cocaine has involved the dopamine system, which will be the focus of this review. Cocaine binds to the dopamine transporter and blocks uptake of dopamine from the synapse. The elevated levels of dopamine activate two superfamilies of dopamine

receptors, D1- and D2-like. This review will focus on D2-like receptors, which have been intimately linked to drug abuse (for instance, see references 10–12). The positron emission tomography (PET) studies that we will describe involve radiotracers that do not differentiate among subtypes of the D2 superfamily (i.e., D_2, D_3, and D_4 receptor subtypes).

To study the brain, powerful imaging techniques have been developed that allow for the noninvasive exploration of brain function in humans and in animals. In this review, we will highlight studies utilizing PET. A clear strength of PET is the ability to examine the brain repeatedly in longitudinal studies. PET has been described as a functional measure of brain activity because it records the uptake and washout of a radioactive marker that competes with endogenous neurotransmitters (13). For example, the radiotracers [^{11}C]raclopride and [^{18}F]fluoroclebopride bind to D2-like receptors with similar affinities. The primary dependent variable in PET imaging studies is the ratio of the distribution of radioligand in the region of interest compared to its distribution in a region devoid of receptors. This ratio, termed the distribution volume ratio, provides a unitless number representing the ratio of receptor density (B_{max}) to receptor affinity. In theory, changes in the distribution volume ratio noted in longitudinal studies reflect changes in B_{max}. However, changes in the distribution volume ratio may also reflect changes in extracellular dopamine levels. Increases in extracellular dopamine will decrease radioligand binding, whereas decreases in extracellular dopamine will increase radioligand binding (for example, see references 13–15). Thus, in addition to being affected by the actual number of receptors in the tissue, the binding of these radioligands is influenced by the amount of dopamine present in the synapse. As a result, when we describe PET data, we use interchangeably "binding" and "availability," since both are being assessed.

In contrast to PET imaging studies, in vitro receptor autoradiography provides a measure of receptor density in a particular brain structure that is not influenced by the levels of a neurotransmitter. The limitation, however, is that the procedures are terminal and, consequently, only one time point can be studied. These are important considerations for longitudinal PET studies because the possibility that changes in PET measures are due to changes in neurotransmitter levels rather than the density of receptors cannot be ruled out unless additional studies are conducted. In this review, we attempt to address this issue by also describing in vitro studies using receptor autoradiography procedures in which circulating dopamine levels are not influencing receptor density assessments.

Animal Models of Drug Abuse: One of the Most Powerful Models of Human Disease

One strategy implemented by the National Institute on Drug Abuse is the development of novel animal models to better understand the neurochemical mechanisms mediating the high abuse potential of cocaine. A greater understanding of the neuropharmacology of cocaine will ultimately lead to improved treatment for cocaine dependence. Animal models of drug self-administration have proven to be valid predictors of human drug abuse (3, 16–20). Animals will self-administer most of the drugs that humans abuse, including cocaine, alcohol, nicotine, and heroin, by the routes (intravenous, inhaled, and oral) used by humans (17). More recently, these models have been extended to examine behavioral and physiological responsiveness to environmental variables as indicators of vulnerability to drug use (21). In one of the earliest studies on individual differences and vulnerability to drug abuse, Piazza et al. (22) initially characterized rats' behavior in an open-field apparatus on the basis of how much locomotor activity was observed prior to any drug exposure. They found that rats characterized as "high responders" in an open field had higher basal corticosterone levels and were more likely to self-administer *d*-amphetamine than rats characterized as "low responders." Consistent with this characterization, noncontingent electric foot shock, which increased corticosterone levels in rats, facilitated acquisition of cocaine self-administration (23). In other experiments, bilateral adrenalectomy or administration of metyrapone, which blocks corticosterone synthesis, completely abolished the acquisition of cocaine self-administration (24). Thus, it appears that individual differences in vulnerability to drug abuse can be observed in animal models and that these differences may be due, in part, to stress-hormone responsiveness.

In addition, animal models of drug abuse have begun to incorporate more sophisticated designs, including the inclusion of alternative reinforcers and the study of cocaine choice, that may provide greater validity than models that simply examine the reinforcing effects of drugs (see references 25–27). For example, when monkeys are given a choice between food and cocaine, there is a dose-dependent relationship between drug dose and preference (see, for instance, references 26–28). That is, when low cocaine doses are the alternative to food, monkeys choose food; at higher cocaine doses, a monkey's preference shifts to almost exclusively cocaine choices. Manipulations of environmental variables can increase or decrease cocaine choice. For example, increases in the magnitude of the nondrug alternative (i.e., food reinforcement) decrease cocaine choice. Similarly, increases in the response requirement for cocaine (analogous to increases in the cost of the drug) decrease cocaine choice (26, 27, 29). The converse is also true: increases in the cost of the preferred nondrug alternative increase the frequency with which cocaine is chosen (26, 27, 29).

These results emphasize that drug reinforcement is not simply mediated by actions in the brain but that the environment can alter drug reinforcement and, as we will show, brain function. A growing body of preclinical re-

search supports the hypothesis that environmental stressors can enhance and environmental enrichment can attenuate the reinforcing effects of drugs. This hypothesis has clear implications for clinical outcomes. As we will discuss, these environmental stimuli affect brain function, including processes mediated by dopamine receptors. Compared to enrichment, more work has focused on the role of stress in drug abuse (see reference 30), with stressors including foot shock or more ethologically relevant variables such as social stress and social defeat. We will describe studies that utilize a novel animal model of drug abuse in which male monkeys live in social groups and have access to cocaine daily.

To relate the preclinical findings to a more global goal of translational research, it is important to point out several advantages of the animal models highlighted in this article. First, drugs can have different effects on brain function depending on whether they are administered by the investigator or self-administered by the animal (31, 32). Because we are interested in studying addictive *behavior,* drug self-administration has face validity as an animal model of a human condition. Second, in our studies of nonhuman primates we also examine complex social behaviors and stable individual behavioral characteristics (e.g., aggressive and affiliative behaviors in particular) that closely model human social interactions (see reference 33). In macaque monkeys, the formation of social hierarchies is determined by the outcome of physical confrontations (i.e., fights). The monkey that wins all of the fights in the social group is the most dominant, the monkey that wins all encounters except with the most dominant animal is the monkey ranked number 2, and so forth. This hierarchy is transitory and linear, such that if number 2 wins a battle with number 3 and number 3 is ranked above number 4, then number 2 is above number 4. We view the linear hierarchy that forms, with the dominant monkeys at the top and the subordinate animals at the bottom, as a continuum from environmental enrichment at one end to high levels of stress at the other. Socially derived stressors have high ethological validity with regard to the study of human drug abuse. Finally, nonhuman primates have many advantages over other animal species. There is abundant evidence that rodent and primate brains differ in the anatomy, physiology, and neurochemistry of brain dopamine systems (34–37). Furthermore, the nonhuman primate brain differs substantially from the rodent brain in terms of cocaine-induced changes in brain metabolism (38, 39). Nonhuman primates can be studied in long-term experiments with cocaine self-administration (over years) that use within-subject designs, and they are capable of learning complex behaviors in order to obtain drugs. The use of PET imaging to examine neuroadaptations allows for longitudinal examination of brain changes due to environmental and/or pharmacological manipulations (e.g., references 40–42). The studies described in the following sections will focus on three phases of drug addiction: vulnerability, maintenance, and abstinence/relapse.

Relation of D2 Receptor Binding to Vulnerability to Cocaine Abuse

Perhaps one of the most challenging issues in drug abuse research involves understanding the etiology of addiction. Epidemiological studies suggest that approximately 17% of the people who use cocaine become dependent on the drug (43, 44). Studies in humans can generate hypotheses about potential brain markers that may indicate a predisposition to become addicted. For example, Volkow et al. (45) studied 23 non-drug-abusing male subjects and used the dopamine transport inhibitor methylphenidate as a tool to study stimulant abuse. First, each subject was scanned with the D2 receptor ligand [^{11}C]raclopride. On another day, they were administered 0.5 mg/kg of methylphenidate and asked to complete an analogue self-rating scale for pleasant drug effects. Approximately one-half (N=12) of the 23 subjects reported liking the dose of methylphenidate, while nine of them described it as "unpleasant" (two were indifferent to the drug effect). Subjects who found methylphenidate "pleasant" had significantly lower levels of D2 receptor binding than subjects who reported the drug as "unpleasant." While these findings suggest an inverse relationship between D2 receptor availability and "vulnerability" to stimulant reinforcement, because of ethical concerns humans cannot be studied prior to drug exposure and then allowed to self-administer the drug for an extended period of time. Animal models, however, can be used to test hypotheses generated from studies of human drug abusers.

We explored further the relationship between D2 receptors and sensitivity to psychostimulants in a novel model of drug abuse in monkeys housed together (40). Previous work from our group had shown a relationship between D2 receptor availability, as measured with PET, and the social rank of female monkeys (46). Subordinate monkeys had significantly lower levels of D2 receptor binding than dominant monkeys. This relationship between social rank and D2 receptors generated an interesting series of questions related not only to drug abuse but also to trait theory. Our first question was whether D2 receptor availability influenced social rank. That is, were monkeys genetically predisposed to a particular position in the social hierarchy according to basal D2 receptor availability? To answer this question, we studied 20 individually housed male monkeys by using the D2 receptor ligand [^{18}F]fluoroclebopride (47) before the monkeys were housed together. The level of D2 receptor binding determined during individual housing did not predict eventual social rank. That is, D2 availability was not a trait variable influencing dominance hierarchies (40).

The next question was whether formation of the social hierarchies influenced D2 receptor availability. When

TABLE 1. Relation of Dopamine D2 Function to Social Rank in Monkeys While Individually Housed and Again After Social Group Formation[a]

	Distribution Volume Ratio (B_{max}/K_d) From PET Study With [^{18}F]Fluoroclebopride					
	Individual Housing		Social Housing		% Change	
Social Rank	Mean	SE	Mean	SE	Mean	SE
Dominant (N=5)	2.49	0.08	3.04	0.23	22.0	8.8
Subordinate (N=5)	2.40	0.06	2.49	0.10	3.9	5.3

[a] The data were originally presented by Morgan et al. (40) in *Nature Neuroscience* (http://www.nature.com/neuro) and are reprinted with permission of Nature Publishing Group.

these animals were rescanned after 3 months of social housing, there were significant differences between groups, with the subordinate monkeys having significantly lower D2 receptor binding than the dominant monkeys (40). This finding replicated our earlier work with female monkeys (46) and extended it to males. However, when we compared each monkey's [^{18}F]fluoroclebopride distribution volume ratio when they were individually housed to the new ratio after social group formation, a profound effect was noted: the average distribution volume ratios for dominant monkeys increased by over 20%, while those for the subordinate animals were nearly unchanged from their original baselines (Table 1). These results suggest that becoming the dominant monkey in the social group produced large changes in dopamine receptor function.

One hypothesis that could account for the observed changes is that being the dominant monkey is analogous to living in an enriched environment. Dominant animals have access to treats in the pen, they are groomed more often than subordinate monkeys, and they move about freely (33). Enrichment could affect the PET signal (i.e., produce increases in the distribution volume ratio) by increasing D2 receptor densities and/or decreasing levels of extracellular dopamine in dominant monkeys. Studies using rodents have clearly shown that environmental enrichment can affect dopamine neurotransmission in an orderly and reliable fashion consistent with both potential mechanisms. For example, Bowling et al. (48)) examined the effects of different rearing environments on dopamine synthesis and metabolism in rats. The group in the "enriched" condition lived together, with 12 or 13 rats per cage, and was exposed to novel toys. Another group lived under "impoverished" conditions consisting of individual housing without toys in the cage. In vitro studies indicated that the rats in the enriched condition had lower striatal concentrations of dopamine than the rats in the impoverished condition. In another study, Hall et al. (49) used in vivo microdialysis to compare rats reared in isolation with socially reared rats, and they found that dopamine levels in the nucleus accumbens were higher in the isolation-reared rats than in the socially reared rats. Furthermore, D2 receptor densities were lower in the isolation-reared rats than in the socially reared rats (49). These findings clearly show that environmental variables as seemingly subtle as housing conditions can produce profound changes in the functional status of dopamine systems. Overall, the findings in rodents are in the same direction as the effects we observed after formation of social groups by monkeys.

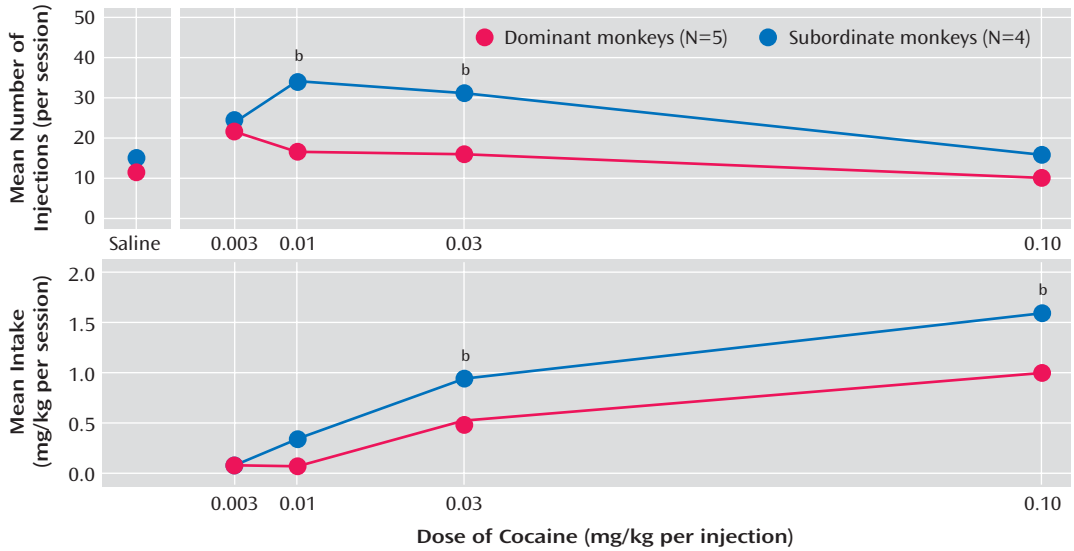
An important question is whether these environmentally induced brain changes have behavioral consequences. In particular, how do these variables influence vulnerability to drug abuse? If, as previously noted in humans, there is a relationship between D2 receptor availability and stimulant reinforcement, then there should be differences in drug reinforcement between dominant and subordinate monkeys and subordinate animals should be more sensitive to the effects of cocaine. This is, in fact, what we found (40). When subsequently allowed access to cocaine, subordinate animals self-administered cocaine at higher rates and had larger intakes than those for dominant monkeys (Figure 1). Overall, these findings confirmed an inverse relationship between D2 receptor availability and cocaine reinforcement and suggested that environmental variables could affect brain function and vulnerability to cocaine abuse.

Decrease of D2 Receptor Binding With Continued Cocaine Use

Human PET imaging studies have shown lower D2 receptor availability in cocaine abusers than in age-matched comparison subjects (11). However, as already described, it is not clear whether the lower D2 receptor binding was a predisposing trait or a consequence of cocaine exposure (see reference 11). Understanding how long-term cocaine exposure affects dopamine receptor function could ultimately lead to better treatment strategies. One hypothesis is that D2 receptor down-regulation occurs as an adaptation to the chronic elevation in extracellular dopamine due to chronic blockade of dopamine uptake. A strength of using animal subjects is the ability to study drug-naive subjects and to observe changes in receptor availability with exposure to a drug by means of a within-subjects, longitudinal design (see reference 50, for example).

We recently extended our studies of socially housed monkeys to examine the effects of chronic cocaine exposure in dominant and subordinate monkeys (51). Whereas dominant monkeys were initially protected from the reinforcing effects of cocaine by elevated D2 availability, chronic exposure to self-administered cocaine resulted in D2 measures that were no longer different from those of subordinate monkeys (Figure 2). These findings suggested that exposure to cocaine attenuated or reversed the powerful effects of environmental context on dopamine receptor availability. Studies utilizing in vitro receptor autoradiography of D2 receptors have consistently shown lower receptor densities in monkeys with long-term his-

FIGURE 1. Relationship of Cocaine Self-Administration to Social Rank in Monkeys[a]



[a] The values are averages from the last 3 days that each dose was available for self-administration. The data are adapted from an article by Morgan et al. (40) in *Nature Neuroscience* (http://www.nature.com/neuro) and are reprinted with permission of Nature Publishing Group.
[b] Significantly different from value for dominant group ($p<0.05$).

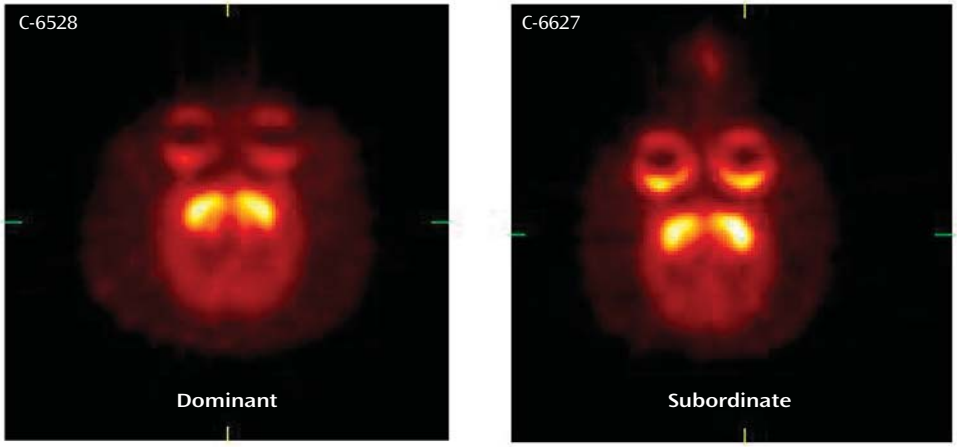
tories of cocaine self-administration; many of these effects were directly related to dose and duration of exposure (52–54).

A central issue for this phase of addiction is one that was raised for the first phase, namely individual differences. For example, one issue to consider in addition to the dose of a drug taken in the lifetime of the individual is the pattern and duration of drug use. Volkow et al. (11) noted that there was not a relationship between D2 receptor binding and the dose of cocaine used, but there was a significant correlation between D2 binding and duration of cocaine use. The effects of duration and cocaine intake can be addressed with animal models in which cocaine availability is the primary independent variable. The question could be phrased as, Does administering a large amount of cocaine over a short period of time produce greater long-term effects on dopamine receptor function than does administering moderate doses over longer periods? In a study not utilizing cocaine self-administration, in vitro receptor autoradiography was used to assess the consequences of chronic cocaine use and abstinence on D2 receptor levels (55). Monkeys were treated four times per day with cocaine for 2 consecutive weeks, followed by a 2-week withdrawal period. The investigators found no differences between cocaine-treated and control monkeys in D2 receptor densities in the caudate nucleus, prefrontal cortex, substantia nigra, and nucleus accumbens. However, this amount of cocaine administered has since been shown to produce robust decreases in D2 receptor densities when *self-administered* over longer periods of time (see, for instance, references 52 and 53). Such findings suggest that the duration of exposure is as important as the actual dose of drug administered and reinforce the importance of using contingent drug administration in animal models of addiction. Future studies using in vivo imaging techniques will better address this extremely important question.

As we discussed for the vulnerability phase, an important question is whether there are functional consequences to cocaine-induced changes in D2 receptor binding. Of particular interest are issues related to cognitive function, decision making, and choice behavior. For example, Grant et al. (56) compared a group of drug abusers (opiate and/or stimulant) and non-drug-abusing comparison subjects on a series of neuropsychological tests to examine the long-term consequences of drug use on decision making. In one set of studies, subjects were exposed to the "gambling task" (57), which has strong face validity for evaluating cognitive deficits (56). The task involved four different decks of cards that differed along three dimensions: immediate gain, long-term expected gain, and schedule of penalties. Two of the four decks had smaller gains and punishers but higher overall "yields"; the comparison subjects chose from those decks most often. In contrast, the majority of the drug abusers chose from the low-yield decks, indicating poor decision making. To control for possible differences in IQ, Grant et al. (56) also used the Wisconsin Card Sorting Test and found no differences between the groups. The authors concluded that impairments in decision making could certainly account for continued use of drugs in the presence of adverse social and personal consequences.

Studies with cocaine abusers cannot clarify whether such cognitive impairments are a consequence of excessive cocaine use or reflect preexisting decrements that predispose certain individuals to drug abuse. Animal studies

FIGURE 2. PET Images of [^{18}F]Fluoroclebopride in Dominant and Subordinate Monkeys With Extensive Histories of Cocaine Self-Administration[a]



[a] The images were coregistered with MRIs from the same monkeys and are shown at the level of the striatum (caudate and putamen). From experiments described by Czoty et al. (51).

can address this issue. Using our model with group-housed monkeys, we studied choice behavior involving cocaine and a nondrug reinforcer—banana-flavored pellets (27). Subordinate monkeys were more sensitive to the reinforcing effects of cocaine than were dominant monkeys. That is, subordinate monkeys chose cocaine over food at a lower cocaine dose than that for dominant monkeys. Such findings suggest that environmental context can substantially influence the preference to choose drug over nondrug alternatives even when D2 binding does not differ between groups.

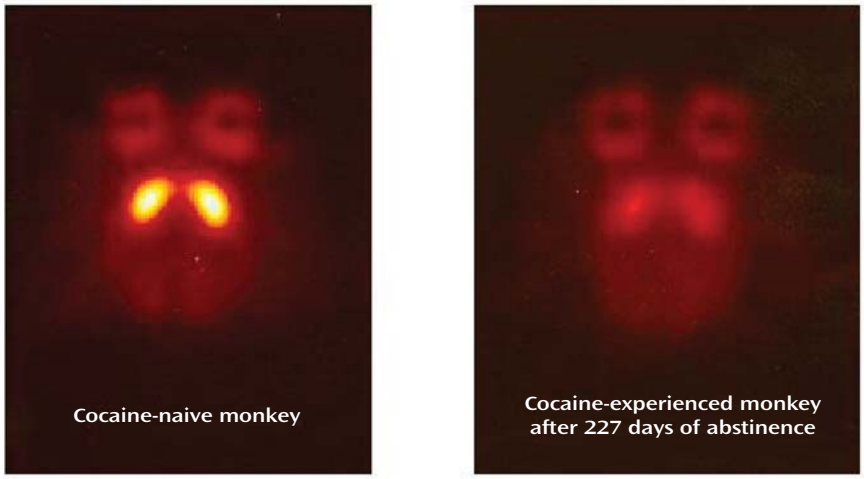
Variable Recovery of D2 Receptor Binding During Cocaine Abstinence

The final phase of drug addiction to be discussed is abstinence and the variables that influence relapse. As with vulnerability and maintenance, we will focus on dopamine D2 receptor availability. From a clinician's viewpoint, abstinence is the critical phase because the population seeking professional guidance and treatment are in this phase; it is worth pointing out that this is the phase of drug addiction about which we know and understand the least. The preceding issues related to vulnerability (predisposition) and maintenance (i.e., changes in brain function due to chronic drug exposure) are critical in understanding brain changes associated with abstinence. However, understanding these variables may provide only a glimpse of the complexity involved in brain changes during abstinence. Perhaps Childress et al. stated it best: "This search [for a cocaine medication] has been complicated by the heterogenous target: drug desire that emerges during cocaine cessation…may well have a different brain substrate than desire induced by cocaine itself…and the cues that signal it" (58, p. 11).

Over a decade ago, Volkow and colleagues (59) reported significantly lower D2 receptor binding in cocaine abusers abstinent from cocaine for up to 4 months. These reductions in D2 receptor binding were associated with decreased glucose utilization in the orbitofrontal cortex and with self-ratings of dysphoria (11). Thus, there was an association between dopamine receptor availability and negative subjective effects during abstinence. The study ended 4 months after the start of abstinence because all subjects had relapsed. Animal models, on the other hand, can provide valuable insight into the neuropharmacology of abstinence and relapse because abstinence can be studied for years.

Our first efforts to study D2 receptor function during abstinence involved using PET imaging to compare D2 receptor binding in monkeys with extensive histories of cocaine self-administration and cocaine-naive comparison monkeys (i.e., a group design). For example, the first monkey we studied had self-administered cocaine for over 3 years and had a lifetime cocaine intake of approximately 20 g. PET scans were conducted at various time points during abstinence. Differences in [^{18}F]fluoroclebopride binding between this cocaine-experienced monkey and a cocaine-naive monkey were apparent at all time points and, as shown in Figure 3, did not dissipate even after 7 months of abstinence. The image from a comparison monkey shows excellent uptake of [^{18}F]fluoroclebopride into D2-rich regions of the basal ganglia. In the abstinent cocaine-experienced monkey, D2 receptor binding was more than 20% lower. While it is possible that the lower D2 signal in the cocaine-experienced monkey was due to elevated levels of dopamine, it is unlikely that such an effect would persist for 7 months. The other possibility is that chronic cocaine exposure resulted in persistent reductions in D2 receptor densities. This hypothesis has been

FIGURE 3. PET Images of [18F]Fluoroclebopride in Cocaine-Naive and Cocaine-Experienced Monkeys[a]



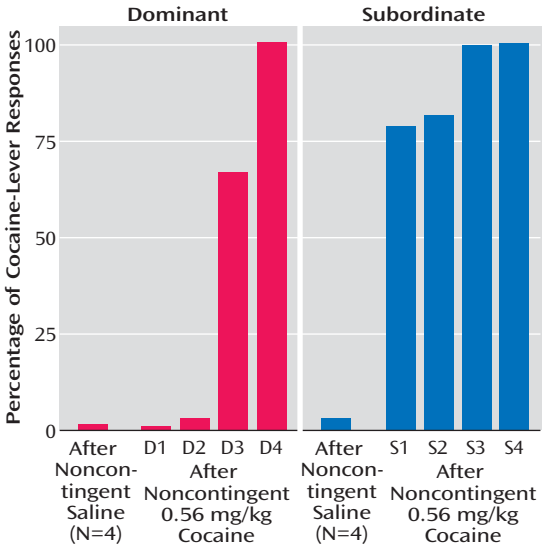
[a] Images are shown at the level of the basal ganglia. From an unpublished study by M.A. Nader and R.H. Mach.

confirmed by means of in vitro receptor autoradiography (52, 53; see reference 54 for further discussion).

It is necessary to comment on animal models and "craving." As defined by Hommer, craving "is a term derived from popular psychology that is used to describe one of the mental states—namely, the intense desire for a certain object or experience" (60, p. 187). Assessing a mental state in an animal is not a realistic undertaking. However, we can measure behavior and hypothesize about mental states. Describing the multitude of animal models of craving is beyond the scope of this review. In brief, the strategies can involve persistence of drug seeking in the presence of stimuli that signal no drug is available (61), escalation of drug seeking with changes in drug availability (62), responding in the presence of conditioned stimuli (63), or a return of responding that had been previously extinguished (i.e., reinstatement) (64).

The reinstatement procedure has been widely used as a model of the ability of priming injections of drugs, drug-related cues, and stress to precipitate relapse into drug taking (reviewed in reference 65). Although the predictive and construct validity of the procedure remain to be firmly established (66), the ability of noncontingent administration of cocaine to reinstate extinguished responding previously maintained by cocaine has been well established (65). Furthermore, it has been demonstrated that drugs that directly or indirectly activate the dopamine system can reinstate cocaine seeking under a food-drug choice procedure (67). Thus, the reinstatement procedure allowed us another means to assess potential differences in dopamine receptor function between dominant and subordinate monkeys. Socially housed monkeys were allowed to self-administer cocaine under conditions in which they also could choose food reinforcement. Under these conditions, as the cocaine dose increased, the percentage of total responses involving the cocaine-associated lever increased (27). When saline was substituted for cocaine, the

FIGURE 4. Effect of a Noncontingent Cocaine Injection on Individual Dominant (D1–D4) and Subordinate (S1–S4) Monkeys' Responses on a Lever Previously Associated With Cocaine but Currently Producing Saline Injections[a]



[a] The first bar for each group represents the average injection-lever responding following noncontingent saline injection. A significant difference was observed after noncontingent injection of 0.56 mg/kg cocaine between the dominant and subordinate monkeys. The data are from an unpublished study by P.W. Czoty, C. McCabe, and M.A. Nader.

monkeys shifted their responses to the food-associated lever (Figure 4). We next examined the ability of noncontingent cocaine to reinstate responses on the cocaine-associated lever, although these responses continued to produce saline injections. Noncontingent cocaine produced dose-related increases in responses on the cocaine-associated lever at doses up to 0.30 mg/kg in all monkeys (data not shown). However, dominant and subordinate monkeys differed in their responses to a higher cocaine dose (0.56 mg/kg). All four subordinate monkeys responded prima-

rily on the injection lever following this dose of cocaine (Figure 4). Greater variability was observed in the dominant monkeys' responses to this dose. These data are consistent with our earlier results demonstrating that subordinate monkeys are, on average, more sensitive to the reinforcing strength of cocaine and/or less sensitive to the aversive effects of cocaine than are dominant monkeys.

Conclusions

In this review we have attempted to briefly highlight the use of animal models and brain imaging procedures (PET and in vitro receptor autoradiography) to convey a better understanding of the neuropharmacology of drug addiction. The focus on dopamine D2 receptors can be viewed as a strength as well as a limitation. The strength of this focus is the understanding of how the level of this receptor superfamily serves as a trait influencing vulnerability to drug abuse and how it serves as a potential state variable showing malleability as a result of environmental or pharmacological manipulations. The weakness of this focus on D2 receptors is the fact that drug addiction is not mediated by one receptor subtype. Future studies must examine how other neurotransmitter and neurohormone systems change in concert with brain dopamine systems, as well as the development of novel radioligands for specific receptor subtypes (e.g., D_3 receptors). In addition, increased anatomical selectivity with higher-resolution PET cameras will better address the contribution of particular brain circuitry to behavioral effects.

The use of animal models to study human disease is an extremely important tool. We have described models that could be characterized as predictive, isomorphic, and homologous of the human condition. In particular, the use of socially housed nonhuman primates to better understand the neuropharmacology and the behavioral and social consequences of drug abuse and to evaluate potential treatment strategies has unequivocal potential. It is our hope that clinicians reading this review will gain a better understanding of the research questions being asked and the potential importance of the results for the development of therapies (behavioral and pharmacological) for drug addiction.

Finally, this review has provided evidence that the environment can profoundly affect drug abuse vulnerability, maintenance, and relapse. Environmental stress and enrichment can influence brain function and vulnerability to drug abuse. Furthermore, even following long-term drug use, environmental variables can affect the behavioral effects of cocaine. The research described in this review represents the growing number of studies documenting the benefits of environmental enrichment, irrespective of genetic predispositions to abuse drugs. Such preclinical findings suggest that outcome measures will be enhanced not only by taking individuals out of a stressful environment but also by providing alternative reinforcers—whether these are better living conditions, jobs, or other activities (68).

Presented at the 157th annual meeting of the American Psychiatric Association, New York, May 1–6, 2004. Received Sept. 15, 2004; revision received Feb. 14, 2005; accepted Feb. 22, 2005. From the Center for the Neurobiological Investigation of Drug Abuse, Department of Physiology and Pharmacology and Department of Radiology, Wake Forest University School of Medicine. Address correspondence and reprint requests to Dr. Nader, Department of Physiology and Pharmacology, Wake Forest University School of Medicine, Winston-Salem, NC 27157; mnader@wfubmc.edu (e-mail).

Supported by National Institute on Drug Abuse grants DA-10584, DA-14637, DA-09085, and P50 DA-06634.

References

1. 2003 National Survey on Drug Use and Health (NSDUH). Rockville, Md, US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, National Institute on Drug Abuse, 2004
2. Emergency Department Trends From the Drug Abuse Warning Network, Final Estimates 1994–2001. Rockville, Md, US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, National Institute on Drug Abuse, 2002
3. Mello NK, Negus SS: Preclinical evaluation of pharmacotherapies for treatment of cocaine and opioid abuse using drug self-administration procedures. Neuropsychopharmacology 1996; 14:375–424
4. Platt DM, Rowlett JK, Spealman RD: Behavioral effects of cocaine and dopaminergic strategies for preclinical medication development. Psychopharmacology (Berl) 2002; 163:265–282
5. Leshner AI: Addiction is a brain disease, and it matters. Science 1997; 278:45–47
6. McLellan AT: Crime and punishment and treatment: latest findings in the treatment of drug-related offenders (editorial). J Subst Abuse Treat 2003; 25:187–188
7. World Health Organization: Neuroscience of Psychoactive Substance Use and Dependence. Geneva, WHO, 2004
8. Madras BK, Fahey MA, Bergman J, Canfield DR, Spealman RD: Effects of cocaine and related drugs in nonhuman primates, I: [^{3}H]cocaine binding sites in caudate-putamen. J Pharmacol Exp Ther 1989; 251:131–141
9. Ritz MC, Lamb RJ, Goldberg SR, Kuhar MJ: Cocaine receptors on dopamine transporters are related to self-administration of cocaine. Science 1987; 237:1219–1223
10. Blum K, Cull JC, Braverman ER, Comings DE: Reward deficiency syndrome. Am Sci 1996; 84:132–145
11. Volkow ND, Fowler JS, Wang GJ, Hitzemann R, Logan J, Schlyer DJ, Dewey SL, Wolf AP: Decreased dopamine D2 receptor availability is associated with reduced frontal metabolism in cocaine abusers. Synapse 1993; 14:169–177
12. Young RM, Lawford BR, Nutting A, Noble EP: Advances in molecular genetics and the prevention and treatment of substance misuse: implications of association studies of the A1 allele of the D2 dopamine receptor gene. Addict Behav 2004; 29:1275–1294
13. Laruelle M: Imaging synaptic neurotransmission with in vivo binding competition techniques: a critical review. J Cereb Blood Flow Metab 2000; 20:423–451
14. Dewey SL, Smith GS, Logan J, Brodie JD, Yu DW, Ferrieri RA, King PT, MacGregor RR, Martin TP, Wolf AP, Volkow ND, Fowler JS, Meller E: GABAergic inhibition of endogenous dopamine release measured in vivo with 11C-raclopride and positron emission tomography. J Neurosci 1992; 12:3773–3780

15. Mach RH, Nader MA, Ehrenkaufer RL, Line SW, Smith CR, Gage HD, Morton TE: Use of positron emission tomography to study the dynamics of psychostimulant-induced dopamine release. Pharmacol Biochem Behav 1997; 57:477–486
16. Spealman RD, Goldberg SR: Drug self-administration by laboratory animals: control by schedules of reinforcement. Annu Rev Pharmacol Toxicol 1978; 18:313–339
17. Griffiths RR, Bigelow GE, Henningfield JE: Similarities in animal and human drug-taking behavior, in Advances in Substance Abuse, vol 1. Edited by Mello NK. Greenwich, Conn, JAI Press, 1980, pp 1–90
18. Johanson CE, Fischman MW: The pharmacology of cocaine related to its abuse. Pharmacol Rev 1989; 41:3–52
19. Woolverton WL, Nader MA: Experimental evaluation of the reinforcing effects of drugs, in Testing and Evaluation of Drugs of Abuse. Edited by Adler MW, Cowan A. New York, Wiley-Liss, 1990, pp 165–192
20. Koob GF, Sanna PP, Bloom FE: Neuroscience of addiction. Neuron 1998; 21:467–476
21. Koob GF, Le Moal M: Drug addiction, dysregulation of reward, and allostasis. Neuropsychopharmacology 2001; 24:97–129
22. Piazza PV, Deminiere JM, Le Moal M, Simon H: Factors that predict individual vulnerability to amphetamine self-administration. Science 1989; 245:1511–1513
23. Goeders NE, Guerin GF: Non-contingent electric footshock facilitates the acquisition of intravenous cocaine self-administration in rats. Psychopharmacology 1994; 114:63–70
24. Goeders NE, Guerin GF: Effects of surgical and pharmacological adrenalectomy on the initiation and maintenance of intravenous cocaine self-administration in rats. Brain Res 1996; 722: 145–152
25. Paronis CA, Gasior M, Bergman J: Effects of cocaine under concurrent fixed ratio schedules of food and IV drug availability: a novel choice procedure in monkeys. Psychopharmacology (Berl) 2002; 163:283–291
26. Negus SS: Rapid assessment of choice between cocaine and food in rhesus monkeys: effects of environmental manipulations and treatment with *d*-amphetamine and flupenthixol. Neuropsychopharmacology 2003; 28:919–931
27. Czoty PW, McCabe C, Nader MA: Assessment of the relative reinforcing strength of cocaine in socially housed monkeys using a choice procedure. J Pharmacol Exp Ther 2005; 312:96–102
28. Nader MA, Wolverton WL: Effects of increasing the magnitude of an alternative reinforcer on drug choice in a discrete-trials choice procedure. Psychopharmacology (Berl) 1991; 105:169–174
29. Nader MA, Woolverton WL: Effects of increasing response requirement on choice between cocaine and food in rhesus monkeys. Psychopharmacology (Berl) 1992; 108:295–300
30. Goeders NE: Stress, the hypothalamic-pituitary-adrenal axis, and vulnerability to drug abuse. NIDA Res Monogr 1998; 169: 83–104
31. Dworkin SI, Mirkis S, Smith JE: Response-dependent versus response-independent presentation of cocaine: differences in the lethal effects of the drug. Psychopharmacology (Berl) 1995; 117:262–266
32. Bradberry CW: Acute and chronic dopamine dynamics in a nonhuman primate model of recreational cocaine abuse. J Neurosci 2000; 20:7109–7115
33. Kaplan JR, Manuck SB, Clarkson TB, Lusso FM, Taub DM: Social status, environment, and atherosclerosis in cynomolgus monkeys. Arteriosclerosis 1982; 2:359–368
34. Berger B, Gaspar P, Verney C: Dopaminergic innervation of the cerebral cortex: unexpected differences between rodents and primates. Trends Neurosci 1991; 14:21–27
35. Haber SN, McFarland NR: The concept of the ventral striatum in nonhuman primates. Ann NY Acad Sci 1999; 877:33–48
36. Joel D, Weiner I: The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. Neuroscience 2000; 96:451–474
37. Cragg SJ, Hille CJ, Greenfield SA: Dopamine release and uptake dynamics within nonhuman primate striatum in vitro. J Neurosci 2000; 20:8209–8217
38. Lyons D, Friedman DP, Nader MA, Porrino LJ: Cocaine alters cerebral metabolism within the ventral striatum and limbic cortex of monkeys. J Neurosci 1996; 16:1230–1238
39. Porrino LJ, Lyons D, Miller MD, Smith HR, Friedman DP, Daunais JB, Nader MA: Metabolic mapping of the effects of cocaine during the initial phases of self-administration in the nonhuman primate. J Neurosci 2002; 22:7687–7694
40. Morgan D, Grant KA, Gage HD, Mach RH, Kaplan JR, Prioleau O, Nader SH, Buchheimer N, Ehrenkaufer RL, Nader MA: Social dominance in monkeys: dopamine D_2 receptors and cocaine self-administration. Nat Neurosci 2002; 5:169–174
41. Howell LL, Hoffman JM, Votak JR, Landrum AM, Wilcox KM, Lindsey KP: Cocaine-induced brain activation determined by positron emission tomography neuroimaging in conscious rhesus monkeys. Psychopharmacology (Berl) 2002; 159:154–160
42. Lindsey KP, Wilcox KM, Votak JR, Goodman MM, Plisson C, Carroll FI, Rice KC, Howell LL: Effects of dopamine transporter inhibitors on cocaine self-administration in rhesus monkeys: relationship to transporter occupancy determined by positron emission tomography neuroimaging. J Pharmacol Exp Ther 2004; 309:959–969
43. Anthony JC, Warner LA, Kessler RC: Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: basic findings from the National Comorbidity Survey. Exp Clin Psychopharmacol 1994; 2:244–268
44. Wagner FA, Anthony JC: From first drug use to drug dependence: developmental periods of risk for dependence upon marijuana, cocaine, and alcohol. Neuropsychopharmacology 2002; 26:479–488
45. Volkow ND, Wang G-J, Fowler JS, Logan J, Gatley SJ, Gifford A, Hitzemann R, Ding Y-S, Pappas N: Prediction of reinforcing responses to psychostimulants in humans by brain dopamine D_2 receptor levels. Am J Psychiatry 1999; 156:1440–1443
46. Grant KA, Shively CA, Nader MA, Ehrenkaufer RL, Line SW, Morton TE, Gage HD, Mach RH: The effect of social status on striatal dopamine D_2 receptor binding characteristics in cynomolgus monkeys assessed with positron emission tomography. Synapse 1998; 29:80–83
47. Mach RH, Luedtke RR, Unsworth CD, Boundy VA, Nowak PA, Scripko JG, Elder ST, Jackson JR, Hoffman PL, Evora PH, Rao AV, Moilinoff PB, Childers SR, Ehrenkaufer RLE: ^{18}F-Labeled radioligands for studying the dopamine D_2 receptor with positron emission tomography. J Med Chem 1993; 36:3707–3720
48. Bowling SL, Rowlett JK, Bardo MT: The effect of environmental enrichment on amphetamine-stimulated locomotor activity, dopamine synthesis and dopamine release. Neuropharmacology 1993; 32:885–893
49. Hall FS, Wilkinson LS, Humby T, Inglis W, Kendall DA, Marsden CA, Robbins TW: Isolation rearing in rats: pre- and postsynaptic changes in striatal dopaminergic systems. Pharmacol Biochem Behav 1998; 59:859–872
50. Thanos PK, Volkow ND, Freimuth P, Umegaki H, Ikari H, Roth G, Ingram DK, Hitzemann R: Overexpression of dopamine D2 receptors reduces alcohol self-administration. J Neurochem 2001; 78:1094–1103
51. Czoty PW, Morgan D, Shannon EA, Gage HD, Nader MA: Characterization of dopamine D1 and D2 receptor function in socially housed cynomolgus monkeys self-administering cocaine. Psychopharmacology (Berl) 2004; 174:381–388

52. Moore RJ, Vinsant SL, Nader MA, Porrino LJ, Friedman DP: Effect of cocaine self-administration on dopamine D_2 receptors in rhesus monkeys. Synapse 1998; 30:88–96
53. Nader MA, Daunais JB, Moore T, Nader SH, Moore RJ, Smith HR, Friedman DP, Porrino LJ: Effects of cocaine self-administration on striatal dopamine systems in rhesus monkeys: initial and chronic exposure. Neuropsychopharmacology 2002; 27:35–46
54. Porrino LJ, Lyons D, Letchworth SR, Freedland CS, Nader MA: Structural and functional neuroimaging of the effects of cocaine in human and nonhuman primates, in Handbook of Neurotoxicology, vol 2. Edited by Massaro EJ. Totowa, NJ, Humana Press, 2002, pp 413–435
55. Farfel GM, Kleven MS, Woolverton WL, Seiden LS, Perry BD: Effects of repeated injections of cocaine on catecholamine receptor binding sites, dopamine transporter binding sites and behavior in rhesus monkeys. Brain Res 1992; 578:235–243
56. Grant S, Contoreggi C, London ED: Drug abusers show impaired performance in a laboratory test of decision making. Neuropsychologia 2000; 38:1180–1187
57. Bechara A, Damasio AR, Damasio H, Anderson SW: Insensitivity to future consequences following damage to human prefrontal cortex. Cognition 1994; 50:7–15
58. Childress AR, Mozley PD, McElgin W, Fitzgerald J, Reivich M, O'Brien CP: Limbic activation during cue-induced cocaine craving. Am J Psychiatry 1999; 156:11–18
59. Volkow ND, Fowler JS, Wolf AP, Schlyer D, Shiue C-Y, Alpert R, Dewey SL, Logan J, Bendriem B, Christman D, Hitzemann R, Henn F: Effects of chronic cocaine abuse on postsynaptic dopamine receptors. Am J Psychiatry 1990; 147:719–724
60. Hommer DM: Functional imaging of craving. Alcohol Res Health 1999; 23:187–196
61. Deroche-Gamonet V, Belin D, Piazza PV: Evidence for addiction-like behavior in the rat. Science 2004; 305:1014–1017
62. Ahmed SH, Koob GF: Transition from moderate to excessive drug intake: change in hedonic set point. Science 1998; 282: 298–300
63. Vanderschuren LJMJ, Everitt BJ: Drug seeking becomes compulsive after prolonged cocaine self-administration. Science 2004; 305:1017–1019
64. de Wit H, Stewart J: Reinstatement of cocaine-reinforced responding in the rat. Psychopharmacology (Berl) 1981; 75:134–143
65. Shaham Y, Shalev U, Lu L, De Wit H, Stewart J: The reinstatement model of drug relapse: history, methodology and major findings. Psychopharmacology (Berl) 2003; 168:3–20
66. Katz JL, Higgins ST: The validity of the reinstatement model of craving and relapse to drug use. Psychopharmacology (Berl) 2003; 168:21–30
67. Gasior M, Paronis CA, Bergman J: Modification by dopaminergic drugs of choice behavior under concurrent schedules of intravenous saline and food delivery in monkeys. J Pharmacol Exp Ther 2004; 308:249–259
68. Higgins ST: The influence of alternative reinforcers on cocaine use and abuse: a brief review. Pharmacol Biochem Behav 1997; 57:419–427

Behavioral/Systems/Cognitive

Genetic Influences on Sociability: Heightened Amygdala Reactivity and Event-Related Responses to Positive Social Stimuli in Williams Syndrome

Brian W. Haas,[1]* Debra Mills,[2]* Anna Yam,[3] Fumiko Hoeft,[1] Ursula Bellugi,[3] and Allan Reiss[1]

[1]Center of Interdisciplinary Brain Sciences Research, Stanford University School of Medicine, Palo Alto, California 94305-5795, [2]School of Psychology, Bangor University, Gwynedd LL57 2AS, United Kingdom, and [3]Laboratory for Cognitive Neuroscience, Salk Institute for Biological Studies, La Jolla, California 92037

Williams syndrome (WS) is a genetic disorder caused by a hemizygous microdeletion on chromosome 7q11.23. WS is associated with a compelling neurocognitive profile characterized by relative deficits in visuospatial function, relative strengths in face and language processing, and enhanced drive toward social engagement. We used a combined functional magnetic resonance imaging (fMRI) and event-related potential (ERP) approach to examine the neural basis of social responsiveness in WS participants to two types of social stimuli, negative (fearful) and positive (happy) emotional facial expressions. Here, we report a double dissociation consistent across both methods such that WS participants exhibited heightened amygdala reactivity to positive (happy) social stimuli and absent or attenuated amygdala reactivity to negative (fearful) social stimuli, compared with controls. The fMRI findings indicate that atypical social processing in WS may be rooted in altered development of disparate amygdalar nuclei that subserve different social functions. The ERP findings suggest that abnormal amygdala reactivity in WS may possibly function to increase attention to and encoding of happy expressions and to decrease arousal to fearful expressions. This study provides the first evidence that the genetic deletion associated with WS influences the function of the amygdala to be particularly responsive to socially appetitive stimuli.

Key words: **Williams syndrome; genetics; fMRI; emotions; amygdala; ERP**

Introduction

Studying the neural basis of social cognition in individuals with well defined genetic etiologies advances our understanding of how genes influence social behavior. Williams syndrome (WS) is a genetic condition caused by a hemizygous microdeletion on chromosome 7q11.23. Individuals with WS are characterized by a distinctive neuropsychological profile comprised of deficient visuospatial function (Meyer-Lindenberg et al., 2004), relative strengths in language (Mervis and Becerra, 2007) and face (Mobbs et al., 2004; Santos et al., 2008) processing, and a tendency for enhanced affiliative drive (Jones et al., 2000; Doyle et al., 2004). For example, those with WS are less socially inhibited (Doyle et al., 2004), more likely to rate emotional facial expressions as approachable (Frigerio et al., 2006), and more gregarious (Klein-Tasman and Mervis, 2003) than mentally aged control groups. Although enhanced drive for social interaction in Williams syndrome has been reported using several experimental (Bellugi et al., 1999; Jones et al., 2000; Mervis and Klein-Tasman, 2000; Klein-Tasman and Mervis, 2003; Deutsch et al., 2007) and cross-cultural (Zitzer-Comfort et al., 2007) approaches, the neural basis of social cognitive function in WS remains poorly understood.

One brain region known to be involved in social functions such as affiliative processes is the amygdala (Skuse et al., 2003). The amygdala codes for the social/emotional salience of both negatively and positively valenced information (Costafreda et al., 2008; Paton et al., 2006), and damage to this structure results in profound abnormalities in social functioning (Aggleton, 2000). Indeed, previous studies have reported functional abnormalities of the amygdala in WS. For example, compared with typically developing controls, WS individuals exhibit diminished amygdala response to fearful stimuli (Meyer-Lindenberg et al., 2005) and heightened amygdala response to music (Levitin et al., 2003). It is unknown, however, if individuals with WS exhibit functional abnormalities when processing positively valenced social/emotional stimuli such as happy facial expressions. Is this study, we tested the hypothesis that those with WS exhibit heightened amygdala response to socially appetitive (happy) stimuli and diminished amygdala response to negative (fearful) social stimuli. We chose to measure amygdala function by means of two complementary methods, functional magnetic resonance imaging (fMRI) and event-related potentials (ERPs), using identical stimuli and a similar experimental design. fMRI provides the opportunity to spatially localize differences in activation during each

Received Nov. 5, 2008; revised Dec. 5, 2008; accepted Dec. 16, 2008.

This work was supported by National Institute of Child Health and Human Development Grant P01 HD033113-12 and National Institute of Mental Health Grant T32 MH19908.

*B.W.H. and D.M. contributed equally to this work.

Correspondence should be addressed to Allan L. Reiss, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, 401 Quarry Road, Stanford, CA 94305-5719. E-mail: areiss1@stanford.edu.

DOI:10.1523/JNEUROSCI.5324-08.2009

Copyright © 2009 Society for Neuroscience 0270-6474/09/291132-08$15.00/0

condition, while ERP provides the opportunity to more accurately dissociate temporal components during each condition. Based on reports that those with WS are relatively hypersocial (Doyle et al., 2004) and have a particular affinity toward happy facial expressions (compared with mentally aged controls) (Frigerio et al., 2006), we predicted that we would observe heightened amygdala response during the processing of happy facial expressions in WS compared with controls. Additionally, based on findings demonstrating that those with WS exhibit reduced amygdala response to fearful social stimuli (Meyer-Lindenberg et al., 2005), we predicted that we would observe a similar pattern as assessed by both fMRI and ERP.

Materials and Methods

Participants. Twenty-seven subjects [13 typically developing (TD), 5 males; and 14 WS, 7 males] participated in the fMRI study, and seventy subjects [25 TD, 11 males; 30 WS, 11 males; and 15, 8 males with developmental delay (DD) of unspecified etiology] participated in the ERP study. Eleven subjects with WS [4 males, mean (M) = 31.13, SD = 7.55] participated in both fMRI scanning and ERP recording. Subjects were excluded from participating if they had a history of psychiatric or neurologic problems as assessed by the Symptom Checklist-90-R (SCL-90-R) (Derogatis, 1977). All included subjects had SCL-90-R scores that fell within one SD of a normative sample. Subjects were also excluded if they reported any current use of mood-altering medication, substance abuse during the 6 months before scan or any standard MRI contraindications.

WS participants were recruited as part of an ongoing multicenter collaborative research study focused on investigating the functional neuroanatomy of WS (33 total; 13 males; mean age = 31.01; SD = 8.80). The diagnosis of WS was genetically confirmed in all participants using the fluorescent *in situ* hybridization test for a deletion of one copy of the elastin gene on chromosome 7. TD subjects were recruited locally (Palo Alto, CA for fMRI and San Diego, CA for ERPs) and were financially compensated for their participation [38 total; 18 males; M = 29.71; SD = 9.52; intelligence quotient (IQ): M = 106, SD = 11.8]. DD participants were recruited locally (San Diego, CA) through community media and state-run agencies (total 15; 8 males; M = 28.07, SD = 8.8; IQ: M = 62, SD = 8.8). Criteria for participation in the DD group included developmental delay of unknown etiology without the presence of symptoms indicative of an autism spectrum disorder. This study was approved by the Stanford University and Salk Institute Administrative Panel on Human Subjects in Medical Research

There were no significant differences in age ($F_{(2,67)} = 0.600$, $p = 0.552$) or proportion of males to females between the groups ($\chi^2_{(2)} = 1.155$, $N = 70$, $p = 0.561$). The WS and DD groups were matched on full-scale IQ (WS: M = 65, SD = 6.8; DD: M = 62, SD = 8.8; $t_{(43)} = 1.148$, $p = 0.153$).

Task design. The stimuli consisted of color pictures of headshots of young adults displaying happy, fearful, and neutral expressions. One hundred undergraduate students were trained to display emotional expressions depicting a variety of emotional expressions including happy, fearful, and neutral. Each photograph was rated by 20 students on a five-point Likert scale for how typical each photograph depicted each emotional category with 1 scored as "not at all like the emotion" and 5 scored as "very characteristic of the emotion." Only stimuli that had the highest average ratings for a given target emotion were selected for that category. Fearful face stimuli were rated as more fearful than neutral ($t = 16.01$, $p < 0.001$) and happy ($t = 18.65$, $p < 0.001$) faces, neutral face stimuli were rated as more neutral than fearful ($t = 36.63$, $p < 0.001$) and happy ($t = 47.54$, $p = 0.001$) faces, and happy face stimuli were rated as more happy than fearful ($t = 61.93$, $p < 0.001$) and neutral ($t = 49.66$, $p < 0.001$) faces. A group of randomly selected photographs were selected to create scrambled isoluminant images that were divided into 256 parts as in a previous study (Mobbs et al., 2004).

Both the fMRI and ERP tasks were event-related designs with four experimental conditions (happy, fearful, neutral, and scrambled) and a resting baseline. Subjects were instructed to judge if each face was male, female, or scrambled by responding with their right index, middle, or ring finger as quickly and as accurately as possible. In the fMRI task, there were a total of 30 trials per condition. Each stimulus was presented for 1750 ms, followed by a 250 ms duration fixation cross. There were two runs, with each run lasting 4 min 32 s. In the ERP task, there were a total of 50 trials per condition. Each stimulus was presented for 1750 ms, followed by a 250 ms duration fixation cross. The intertrial interval was random with a mean of 1000 ms (range = 800–1200 ms). The random order of the intertrial interval was determined by using Presentation software (Neurobehavioral Systems). Behavioral responses were collected within the 2000 ms following the onset of each stimulus in both the fMRI and ERP task.

fMRI data acquisition. Whole-brain imaging data were acquired on a GE-Signa 3 T scanner (General Electric). For structural whole-brain images, a three-dimensional high-resolution spoiled gradient scan (SPGR) (repetition time, 24 ms; echo time, 5 ms; flip angle, 15°; number of excitations, 2; matrix size, 256 × 256; field of view, 24 cm; slice thickness, 1.2 mm; 124 contiguous slices) and a T1 in-plane scan (14 slices, 5 mm thickness; oriented parallel to the line between the anterior and posterior commissure) were conducted. Functional images were acquired using a gradient echo T2*-weighted echoplanar imaging scan and were obtained using a flip angle of 80°, repetition time = 2.0 s, echo time = 30 ms, 32 slices, and a field of view = 200 × 200 mm matrix.

Functional data were preprocessed and statistically analyzed using SPM2 (Wellcome Department of Imaging Neuroscience, London, UK). The images were temporally realigned to the middle slice and spatially realigned to the first in the time series. The images were then coregistered and spatially normalized into standard stereotactic space (MNI template). All images were spatially smoothed with an 8 mm full width-half maximum isotropic Gaussian filter.

ERP data acquisition. The electroencephalogram was recorded using a high-density 64-channel geodesic net from Electrical Geodesics and sampled at 250 Hz, with a bandpass of 0.1–100 Hz. The electrooculogram was recorded from over and under the left eye to monitor blinks and vertical eye movements and from the right outer canthus to monitor horizontal eye movements. Impedances were maintained below 50 kΩ. The averaged ERPs were also digitally filtered off-line with a 30 Hz low-pass filter. All electrodes were referenced to the vertex during recording and rereferenced off-line to the average of the left and right mastoids for analysis. Artifact rejection was conducted off-line using an automatic computer program to reject trials containing blinks and vertical and horizontal eye movements (±70 μV). A mean of 88% of the trials (mean = 44/50 trials per condition, SD = 7.06) for the TD group, 70% (35/50 trials, SD = 12.71) for the WS group, and 64% (32/50 trials, SD = 11.05) for the DD group were retained for analysis. The WS group had fewer artifact-free trials per condition than the TD group main effect of group ($F_{(1,53)} = 10.00$, $p = 0.003$) but did not differ from the DD group ($F_{(1,43)} = 0.601$, $p = 0.443$). Two additional participants with WS, one TD, and one DD had <10 trials for a given condition and were excluded from the analysis.

fMRI data analysis. Fixed-effects models representing two runs for each participant were used at the individual subject level of analysis and random effects models were used for group-level analyses (SPM2). At the individual level, models were created that represented all event-related task conditions (happy, fearful, neutral, scrambled, and fixation). Each stimulus presentation was modeled as a single event. Data were high-pass filtered. Images identified to correspond with >2 mm of motion were not included in the statistical analysis. There was no significant difference in the number of images excluded between groups (TD average = 0.15; WS = 0.50). Images were not excluded if they corresponded with incorrect responses.

Statistical analyses were initiated by performing a condition-by-group (2 × 2) interaction analysis. This was conducted by using a random effects model comparing the difference in blood oxygenation level-dependent (BOLD) signal during each emotional condition (happy or fearful) versus neutral between experimental groups (WS vs TD). The simple effects within each experimental group were then explored by comparing BOLD signal change between each of the emotional conditions (happy or fearful) versus neutral.

Based on previous studies reporting atypical amygdala function in WS (Meyer-Lindenberg et al., 2005) and previous studies reporting changes

in amygdala function during the processing of positive social stimuli (Canli et al., 2002; Skuse et al., 2003; Paton et al., 2006; Costafreda et al., 2008), we designated the amygdala as an a priori region of interest (ROI). The delineation of the left and right amygdala ROIs were based on Talairach definitions in standard stereotactic space (http://www.fmri.wfubmc.edu). Unless otherwise specified, we used a $p < 0.05$ statistical threshold (corrected for multiple comparisons within the amygdala) for this analysis. Event-related responses were extracted from significant clusters within the amygdala using a moving average for each time course of 200 ms. Percentage signal change between conditions was calculated based on the difference of the peak response extracted from each cluster during either happy or fearful conditions compared with neutral.

ERP data analysis. Numerous studies have identified ERP components sensitive to emotional content, including the components measured here: an N200 linked to arousal and attention (Streit et al., 1999; Sato et al., 2001; Balconi and Lucciari, 2007), an anterior positivity linked to valence (Dolcos and Cabeza, 2002), and a posterior positivity linked to arousal and/or increased subsequent memory for emotional stimuli (Diedrich et al., 1997; Dolcos and Cabeza, 2002; Eimer and Holmes, 2007; Koenig and Mecklinger, 2008). Based on the current data and previous studies, ERP emotion effects were measured using mean amplitudes for the N200 between 200 and 280 ms, the anterior positivity from 500 to 700 ms, and the posterior positivity from 300 to 500 ms. A peak latency measure was also taken for the N200 between 200 and 280 ms. As the distributions of these emotion effects varied over the scalp, the electrodes used in the statistical analysis differed for each ERP component (supplemental Fig. 1, available at www.jneurosci.org as supplemental material). Trials were not excluded if they corresponded with incorrect responses.

Group comparisons of the difference waveforms for happy − neutral and fearful − neutral for the N200, P300–500, and P500–700 were analyzed in repeated-measures ANOVAs using SPSS 13 with three levels of group (WS, TD, and DD) as the between-subjects variable, two levels of hemisphere (left and right), and electrode site (number is variable according to the montage) as repeated measures. The happy versus neutral comparison was conducted over central posterior sites, and thus there was no comparison across hemisphere. Planned comparisons for each group analyzed separately were conducted for each component with two levels of emotion (e.g., happy vs neutral, or fearful vs neutral), two levels of hemisphere for fearful versus neutral, and electrode site (same sites as used in the group analyses). The Greenhouse-Geisser correction was used for main effects and interactions with more than one electrode site. To control for differences between men and women in the effects of emotion on brain activity, preliminary analyses included gender as a separate between-subjects factor in the ANOVA. There were no interactions with emotion and gender for any of the groups; therefore, gender was not included as a factor in the analyses reported below.

Results

Behavioral measures

Mean reaction times and accuracy were submitted to a general linear model of repeated measures with subject group (WS, TD, and DD) and experimental condition (happy, fearful, neutral, and scrambled) entered as factors. Overall, WS subjects (mean reaction time: 1222.66) responded slower compared with TD subjects (796.03) ($F = 46.61$, $p < 0.001$) but not DD (1377.66) subjects ($F = 1.35$, $p = 0.25$). Similarly, WS (88.33% accuracy) subjects were less accurate in identifying gender than TD (97.36% accuracy) ($F = 0.15.47$, $p < 0.001$) subjects, but not DD (88.97%) subjects ($F = 0.03$, $p = 0.87$). Participants in each group and in each study (fMRI and ERP) were cognizant of all task instructions and performed the task above a threshold for accuracy of 75%. There was no between-group effect in the difference between reaction time (RT) or accuracy during either of the emotional conditions (RT: happy: $t = -0.73$; $p > 0.10$; fearful: $t = 0.55$; $p > 0.10$; accuracy: happy: $t = -0.56$; $p > 0.10$; fearful: $t = 1.36$; $p > 0.10$) compared with the neutral condition (condition × group interaction).

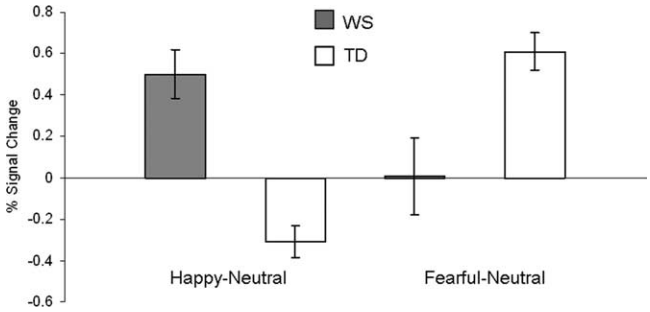


Figure 1. Right amygdala activation to happy (left) and fearful (right) facial expressions compared with neutral between groups. Bars represent percentage signal change of clusters localized by means of direct comparisons between the TD and WS groups. Error bars represent SEM.

Amygdala reactivity to happy facial expressions

We compared happy relative to neutral amygdala response between the two experimental groups (condition × group interaction). Relative to TD participants, the WS group exhibited greater right amygdala reactivity to happy compared with neutral facial expressions ($p < 0.05$ corrected; 28 voxels; MNI: 30, 0, −14) (Fig. 1, left). On the other hand, the TD group did not exhibit significantly greater left or right amygdala reactivity to happy compared with neutral facial expressions than the WS group. We next compared BOLD signal within the amygdala during the response to happy and neutral facial expressions in the WS and TD groups independently (Fig. 2). The WS group exhibited significant right amygdala reactivity to happy compared with neutral facial expressions ($p < 0.05$ corrected; 16 voxels; MNI: 32, −2, −14) (Fig. 2*B*). This cluster encompassed a relatively superior region ($z = -14$) within the standardized amygdala ROI. No significant left or right amygdala activation in response to happy compared with neutral facial expressions was observed in the TD group.

Amygdala reactivity to fearful facial expressions

We compared fearful relative to neutral amygdala response between the two experimental groups (condition × group interaction). Relative to WS participants, the TD group exhibited greater right amygdala reactivity to fearful compared with neutral facial expressions at a level approaching statistical significance ($p = 0.064$ corrected, $p = 0.002$ uncorrected, two voxels; MNI: 20, 0, −14) (Fig. 1, right). To explore the consistency of our findings with that of a previous study (Meyer-Lindenberg et al., 2005), we also compared amygdala response to fearful expressions relative to "nonsocial" stimuli (scrambled faces). Relative to the WS group, the TD group exhibited greater right amygdala reactivity to fearful compared with scrambled faces at the corrected statistical threshold ($p < 0.05$ corrected, 19 voxels; MNI: 24, −2, −16). There was no significant difference in amygdala reactivity to neutral relative to scrambled faces between the WS and TD groups. Additionally, the WS group did not exhibit significantly greater left or right amygdala reactivity to fearful (compared with either neutral or scrambled) facial expressions than the TD group. We next compared BOLD signal within the amygdala during the response to fearful and neutral facial expressions in the TD and WS groups independently (Fig. 2). The TD group exhibited significant right amygdala reactivity to fearful compared

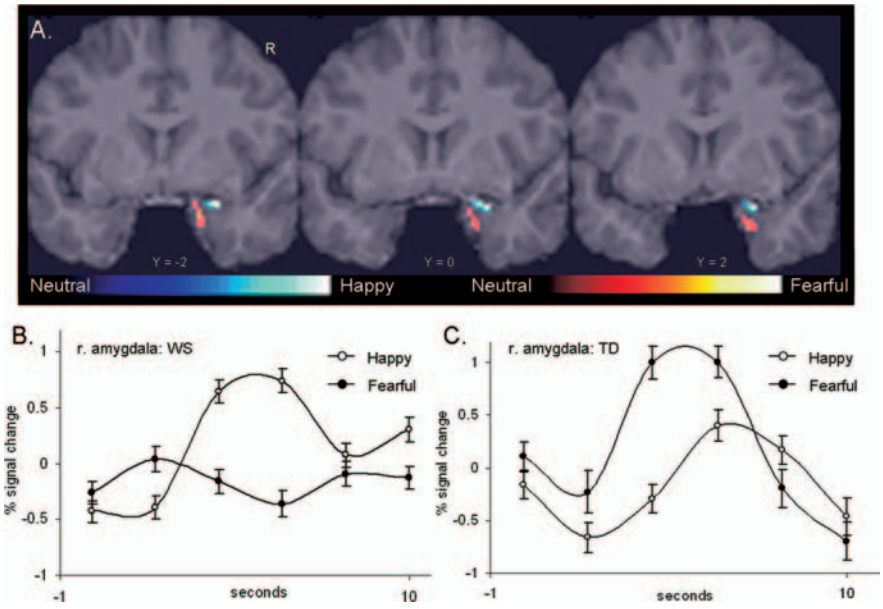


Figure 2. Areas of greater right amygdala reactivity to fearful and happy facial expressions (compared with neutral) within the WS and TD samples. ***A***, These clusters were not overlapping. We reduced the threshold to $p < 0.05$ uncorrected to visualize the distinct location for each cluster. No voxels within the right amygdala were found to be significantly activated in response to fear versus neutral in the WS group or happy versus neutral in the TD group at the same statistical threshold. Clusters are overlaid onto three coronal slices (Montreal Neurological Institute coordinates: $y = -2, 0, 2$) of a representative TD brain normalized into standard stereotactic space. Voxels of greater activation in response to happy versus neutral facial expressions in the WS group are designated by cool colors (blue). Voxels of greater activation in response to fearful versus neutral facial expressions in the TD group are designated by warm colors (orange). Extracted time course of percentage signal change for peak voxels within the clusters displayed in ***A*** are shown in ***B*** for WS participants and in ***C*** for TD participants. Data were extracted and converted to percentage signal change using a moving average (200 s) by a standardized method (xjview toolbox in SPM2). Lines denoted with open circles represent BOLD response to happy faces. Lines denoted with filled circles represent BOLD response to fearful faces. Error bars represent SEM. R, Right.

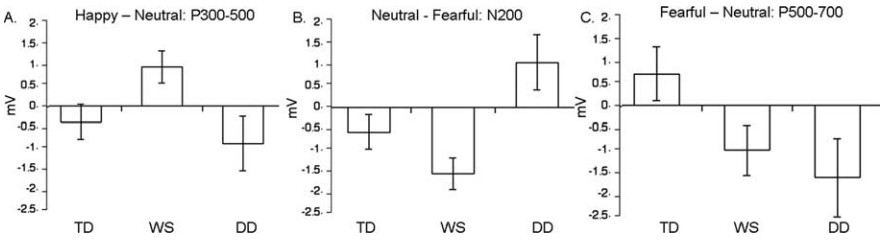


Figure 3. Direct comparison between groups for the P300–500 during happy relative to neutral facial expressions and the N200 and P500–700 during fearful relative to neutral facial expressions. ***A***, The P300–500 difference for happy − neutral was larger for WS than TD or DD groups. ***B***, For the N200, the WS group showed a larger N200 difference to neutral − fearful than TD or DD groups. ***C***, The P500–700 difference to fearful − neutral was larger for TD than WS or DD groups.

with neutral ($p < 0.05$ corrected; 12 voxels; MNI coordinates: 20, −6, −20) (Fig. 2*C*) and scrambled ($p < 0.05$ corrected; 39 voxels; MNI: 22, −4, −22) facial expressions. On the other hand, no significant left or right amygdala activation to fearful compared with neutral or scrambled facial expressions was observed in the WS group.

Amygdala response associated with IQ, reaction time, and accuracy

The fMRI study did not include a developmentally delayed control group as was available in the ERP study. Therefore, within the WS group, we examined the association of amygdala response with cognitive and behavioral variables that differentiated this group from TD controls. First, we entered IQ as a covariate with BOLD signal extracted from the right amygdala and found that there were no significant correlations between IQ and amygdala activity either in response to happy ($p = 0.57$) or fearful ($p = 0.48$) facial expressions within the WS group. Second, we entered reaction time and accuracy as covariates with BOLD signal extracted from the right amygdala and again found that there were no significant correlations between either reaction time or accuracy and activity either in response to happy or fearful facial expressions within the WS group (all p values >0.10).

ERPs

The organization of the ERP results section is designed to mirror the comparisons reported in the fMRI findings above. Thus, comparisons of ERP mean amplitudes linked to happy versus neutral expressions (P300–500) are followed by fearful versus neutral expressions (N200 and P500–700) respectively. In each section, we first report the overall emotion (happy or fearful vs neutral) × group (WS, TD, DD) interaction for a given ERP component. This is followed by tests of the predicted group differences for specific ERP emotion effects by using difference waves (e.g., P500 to happy minus neutral). These a priori hypotheses are further examined for each group individually. Finally, the specificity of the targeted emotion effects is examined for each ERP component. That is, where a significant emotion (happy or fearful vs neutral) × group (WS, TD, DD) interaction is observed, we conduct the same condition × group interaction analysis but substitute happy with fearful or fearful with happy. Throughout the ERP results section, p values that are <0.05 are referred to as significant, 0.05–0.10 as trends, and >0.10 as nonsignificant.

ERPs to happy facial expressions: P300–500

Based on visual inspection of the ERPs and previous research, ERPs to happy versus neutral expressions were examined from 300 to 500 ms (P300–500) for the WS, TD, and DD groups over six posterior sites (EGI locations, left: 29, 33, 42; right: 41, 34, 38) (Fig. 3*A*). The condition × group interaction indicated that the P300–500 response to happy versus neutral facial expressions was significantly different between the three groups ($F_{(2,66)} = 3.61$, $p = 0.03$, $\eta_p^2 = 0.100$) (Fig. 3*A*). We next tested the a priori hypothesis that the WS group would exhibit a greater P300–500 response to happy versus neutral facial expressions than the TD and DD control groups. Between-group comparisons showed that the P300–500 difference to happy minus neutral facial expressions was larger for the WS than for the TD group ($F_{(1,53)} = 7.189$, $p = 0.01$, $\eta_p^2 = 0.119$) or the DD group ($F_{(1,42)} = 5.362$, $p = 0.026$, $\eta_p^2 = 0.113$) but did not differ between the TD and DD groups ($F_{(1,37)} =$

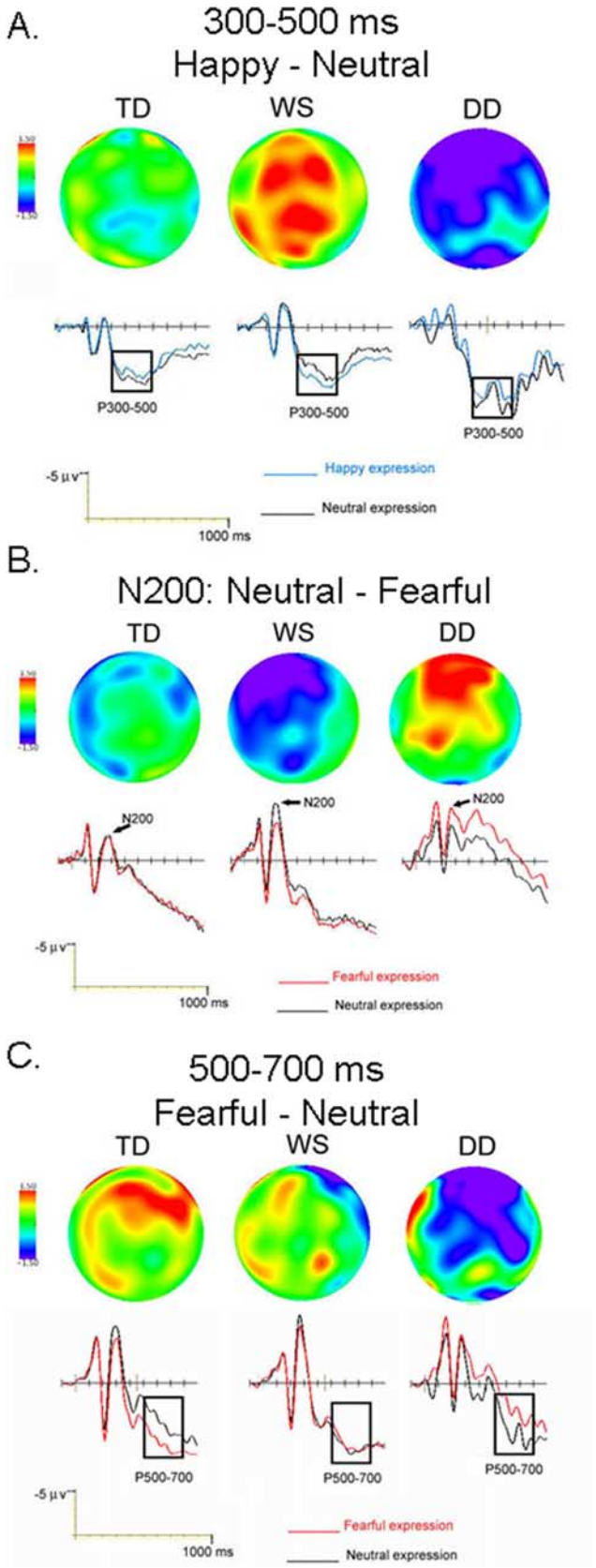


Figure 4. Event-related potentials and topographical maps comparing responses to different emotional expressions for the TD (left), WS (center), and DD (right) groups. Note that negative voltage is plotted up in all ERP plots. Topographical maps (top of each panel) illustrate the distribution of the difference in activity between conditions. Amplitude is plotted as a function of color with the most negative voltage as purple and most positive voltage as red. ***A***, ERPs to happy (blue lines) compared with neutral (black lines) expressions depicted at site 42 (P2). For the WS group, the P300–500 (boxed area) was larger to happy than neutral expressions. The P300–500 did not differ by emotional expression for the other two groups. The posterior central distribution of the happy − neutral difference wave at 350 ms is shown in the topographical map. ***B***, ERPs to neutral (black lines) compared with fearful (red lines) expressions shown at electrode 12 (AF3). The N200 amplitudes to fearful compared with neutral expressions are decreased in the WS group, and increased in the DD group. The topographical maps illustrate the anterior distribution of the N200 difference to fearful − neutral expressions at 240 ms. A larger N200 to fearful than neutral expression results in a positive difference that is represented in red as in the DD group (right side). In contrast, a smaller N200 to fearful than neutral expression results in a difference that is negative in voltage represented in purple as for the WS group. ***C***, ERPs to neutral (black lines) and fearful (red lines) expressions are shown at the bottom half of the figure shown at electrode 57 (next to FC4). The anterior positivity from 500 to 700 ms, P500–700 (boxed area), was larger to fearful than neutral expression for the TD controls (left) and did not differ for the WS group. For the DD group, the P500–700 appears to be larger to neutral than fearful expressions at this site, but the effect was not significant at this or other sites. The anterior distribution of the P500–700 difference wave to fearful − neutral expression at 520 ms is illustrated in the topographical maps.

0.621, $p = 0.436$). We next compared P300–500 response to happy versus neutral facial expressions in the WS, TD, and DD groups independently (Fig. 4*A*). The WS group exhibited a greater P300–500 response to happy versus neutral facial expressions ($F_{(1,29)} = 4.502$, $p = 0.043$, $\eta_p^2 = 0.136$). In contrast, the P300–500 response to happy versus neutral facial expressions was not different for the TD ($F_{(1,24)} = 0.892$, $p = 0.354$) or DD ($F_{(1,14)} = 1.817$, $p = 0.201$) groups.

We examined the specificity of emotion by comparing fearful versus neutral P300–500 response over posterior regions between the three (WS, TD, and DD) experimental groups (condition × group interaction). The condition × group interaction indicated that the P300–500 response to fearful versus neutral facial expressions was not significantly different between the three groups ($F_{(2,66)} = 1.426$, $p = 0.248$), nor was the P300–500 comparison to fearful versus neutral expressions significant for any individual group: TD, $F_{(1,24)} = 0.046$, $p = 0.832$; WS, $F_{(1,29)} = 0.374$, $p = 0.546$; DD, $F_{(1,14)} = 1.435$, $p = 0.252$.

Event-related responses to fearful facial expressions: N200 and P500–700

N200

We compared the N200 mean amplitude and peak latency to fearful versus neutral expressions for the three groups (WS, TD, and DD) over the left and right hemisphere (total of 16 sites) (see supplemental Fig. 1, available at www.jneurosci.org as supplemental material). The condition × group interaction indicated that the N200 to fearful versus neutral facial expressions was significantly different between groups in mean amplitude ($F_{(2,66)} = 4.531$, $p = 0.014$, $\eta_p^2 = 0.122$), group × emotion × electrode site ($F_{(2,66)} = 2.363$, $p = 0.034$, $\eta_p^2 = 0.067$) (Fig. 3*B*). There were no condition × group effects for peak latency ($F_{(2,66)} = 0.290$, $p = 0.749$). We next tested the a priori hypothesis that the WS group would exhibit decreased neural activity to fearful versus neutral facial expressions relative to the TD and DD control groups. Between-group comparisons showed that the N200 difference wave to fearful minus neutral facial expressions tended to be greater (more negative) for the WS than for the TD group over the left hemisphere ($F_{(1,53)} = 3.321$, $p = 0.07$) and was greater (more negative) for WS than for the DD group ($F_{(1,42)} = 8.327$, $p = 0.006$, $\eta_p^2 = 0.165$). In contrast, the effect was smaller (more positive) for the DD than for the TD group ($F_{(1,37)} = 4822$, $p = 0.034$, $\eta_p^2 = 0.115$). We next compared the N200 to fearful versus neutral facial expressions in the WS, TD, and DD groups inde-

pendently (Fig. 4*B*). For the WS group, the N200 was smaller to fearful than neutral expressions over the left hemisphere, condition × hemisphere ($F_{(1,29)} = 13.048$, $p = 0.001$, $\eta_\rho^2 = 0.316$), and peaked significantly earlier ($F_{(1,29)} = 14.350$, $p = 0.001$). In contrast, for the TD group, the N200 did not differ in mean amplitude to fearful versus neutral expressions ($F_{(1,24)} = 1.253$, $p = 0.214$) or peak latency ($F_{(1,23)} = 1.997$, $p = 0.171$). Although the DD group did not show N200 mean amplitude ($F_{(1,14)} = 2.089$, $p = 0.172$) or latency ($F_{(1,14)} = 0.724$, $p = 0.495$) differences over these electrodes, they did show increased N200 amplitudes to fearful relative to neutral expressions over different central (more medial) regions (13, 4, 62; F5, F3, F4) ($F_{(1,14)} = 4.847$, $p = 0.046$, $\eta_\rho^2 = 0.272$).

We examined the specificity of the N200 responsiveness to fearful versus neutral emotions by comparing the N200 amplitude to happy versus neutral expressions over the same 16 regions for the three (WS, TD, and DD) experimental groups (condition × group interaction). The N200 response to happy versus neutral facial expressions was not significantly different between the three groups ($F_{(2,66)} = 2.185$, $p = 0.121$).

P500–700

We compared the P500–700 mean amplitude to fearful versus neutral expressions for the three groups at 14 sites (EGI sites, left: 14, 12, 19, 15, 13, 20, 16; right: 1, 2, 60, 61, 62, 56, 57) (Fig. 3*C*; supplemental Fig. 1, available at www.jneurosci.org as supplemental material). The predicted emotion × group × hemisphere interaction only approached significance ($F_{(2,66)} = 2.838$, $p = 0.066$). Based on the current fMRI findings showing increased right amygdala activation to fearful relative to neutral stimuli in the TD group, and previous ERP research showing an increased positivity over right anterior regions in the 500–700 ms range linked to fearful relative to neutral expressions, we predicted that the TD group would show an increased positivity, whereas the WS group would show a diminished positivity, to fearful relative to neutral expressions. Therefore, we tested the a priori hypothesis that the WS group would exhibit decreased response to fearful versus neutral facial expressions relative to the TD and DD control groups. Between-group comparisons showed that the P500–700 difference to fearful minus neutral facial expressions was larger for the TD group than for the WS groups over right anterior regions ($F_{(1,53)} = 4.337$, $p = 0.042$, $\eta_\rho^2 = 0.076$). The P500–700 difference to fearful minus neutral facial expressions did not significantly differ between the WS and the DD groups ($F_{(1,42)} = 1.717$, $p = 0.197$) at any electrode site. However, there was a trend for the P500–700 difference to fearful minus neutral facial expressions to be different between the TD and DD groups ($F_{(1,37)} = 3.329$, $p = 0.08$). We next compared P500–700 response to fearful versus neutral facial expressions in the WS, TD, and DD groups independently (Fig. 4*C*). For the WS group, the P500–700 response to fearful versus neutral facial expressions was smaller over the right hemisphere ($F_{(1,29)} = 6.376$, $p = 0.017$, $\eta_\rho^2 = 0.180$). For the TD group, the P500–700 response to fearful versus neutral facial expressions tended to be larger over right anterior regions [$F_{(1,24)} = 3.335$, $p = 0.08$, $\eta_\rho^2 = 0.115$; at electrode 57 (FC6), $t_{(24)} = 2.342$, $p = 0.028$]. For the DD group, the P500–700 response to fearful versus neutral facial expressions was not significantly different at any site ($F_{(1,14)} = 1.345$, $p = 0.267$).

We examined the specificity of emotion by comparing happy versus neutral P500–700 response over the same regions between the three (WS, TD, and DD) experimental groups. The emotion × group interactions indicated that the P500–700 response to happy versus neutral facial expressions was not significantly different between the three groups ($F_{(2,66)} = 1.081$, $p = 0.345$; group × emotion × hemisphere, $F_{(2,66)} = 0.152$, $p = 0.860$).

Discussion

Results from the fMRI and ERP studies using identical stimuli and a similar experimental design provided convergent evidence that individuals with WS exhibit both heightened amygdala reactivity to happy facial expressions and diminished amygdala reactivity to fearful facial expressions. The heightened amygdala response to socially appetitive stimuli may be a fundamental neural substrate that underlies atypical social drive in WS. Similar to the results of a previous study (Meyer-Lindenberg et al., 2005), we also found that TD healthy controls exhibited a robust amygdala response to fearful facial expressions, while the WS group did not, although the condition × group interaction (using neutral as a baseline) approached corrected statistical significance. By using an additional methodological approach (ERPs), we were able to provide an electrophysiological correlate of these fMRI findings and to examine the temporal specificity of the neural response to fearful stimuli within each group. Specifically, the WS group showed decreased activity to fearful versus neutral expressions as reflected by decreases in the mean amplitudes of the N200, compared with both the DD and TD groups. The significant between-group differences in N200 amplitude provide evidence that relatively early during the visual and perceptual stream, WS individuals differ from both healthy and IQ-matched controls in how fearful facial expressions are processed.

Of particular interest here was that the WS group showed increased activity to happy relative to neutral expressions as evidenced by relative increases in BOLD signals in the right amygdala and amplitude of the posterior P300–500 waveform. The posterior P300–500 ERP effect has been linked to enhanced memory for emotional content and may reflect increased attention to emotionally salient stimuli during encoding (Deidrich et al., 1997; Dolcos and Cabeza, 2002; Koenig and Mecklinger, 2008). Regarding the ERPs to fearful faces, we did not predict that the P500–700 difference to fearful minus neutral facial expressions would be different between the TD and DD groups. An ERP/fMRI study of positive, negative, and emotional scenes showed that the amplitude of the late posterior positivity indexes emotional intensity (Sabatinelli et al., 2007) and is correlated with activation from the amygdala (Sabatinelli et al., 2005). The present data represent a new finding and are the first to show increased neural activity to positive emotional expressions in WS.

Considering the fMRI findings for both positively and negatively valenced social stimuli, the clusters of fMRI activation within the right amygdala were observed to be nonoverlapping (Fig. 2*A*). This suggests that distinct neuronal populations may subserve these functions and that this effect may be primarily lateralized to the right amygdala. Recently, single-unit recording studies in primates have indicated the presence of emotion selective neurons within the amygdala (Paton et al., 2006; Kuraoka and Nakamura, 2007). Although some lesion studies in rats indicate that disparate nuclei within the amygdala are engaged in different functions (Knapska et al., 2007; Yang et al., 2008), other studies suggest that amygdala nuclei function in parallel during appetitive and aversive processing (Balleine and Killcross, 2006). The fact that the results in this study are based on averaged, spatially smoothed data limits how precisely the clusters can be assigned to specific nuclei within the amygdala. However, a single-subject analysis based on nonsmoothed data confirms the general topographical dissociation of amygdala activation be-

tween groups but also demonstrates the heterogeneity of BOLD signal within groups (supplemental Fig. 2, available at www.jneurosci.org as supplemental material). The putative dissociation of these loci suggests that the genetic deletion in WS may influence the development and function of separate amygdala regions (and possibly nuclei) differently.

The increased amygdala response to happy facial expressions in individuals with WS may represent several psychological processes linked to the amygdala including attention, arousal, or anxiety. For example, the central nucleus of the amygdala has been strongly linked with attention (Holland and Gallagher, 1999; Maddux et al., 2007). Although eye-tracking information would have helped clarify this issue in our study, unfortunately, equipment to obtain these data were not available at the time of subject scanning. However, other investigations have assessed attentional mechanisms during social processing in WS by measuring gaze duration of infants and toddlers toward faces (Mervis et al., 2003), or by performing eye tracking when subjects viewed social and nonsocial stimuli (Riby and Hancock, 2008). The results of these studies indicate that WS is associated with enhanced attention for social and emotional stimuli. For example, whereas individuals with autism tend to exhibit reduced fixation toward faces and eyes, people with WS exhibit heightened amounts of fixation toward faces and eyes (Riby and Hancock, 2008) relative to typically developing and nonverbal ability-matched controls. Although demonstrating a WS-associated predisposition for increased attention to social–emotional stimuli in general, these studies did not directly address differential attention for or neural responses to positive versus negative emotional stimuli in affected individuals.

The social phenotype in WS has been described as being comprised of relative strengths and weaknesses (Mervis, 2003; Feinstein and Singh, 2007). On one hand, those with WS appear to be particularly driven toward engaging in social interaction and to approach others such as strangers excessively and inappropriately. Neuropsychological research has demonstrated that individuals with WS tend to rate happy facial expressions as more approachable relative to other emotions than do mentally aged controls (Frigerio et al., 2006; Porter et al., 2007). On the other hand, those with WS often experience profound difficulties maintaining relationships and also experience heightened levels of anxiety during social interaction (Dykens, 2003). The data from the current study is consistent with this behavioral phenotype of WS and provides neurobiological evidence that the WS social phenotype is mediated by both increased neural reactivity to happy expressions and decreased reactivity to fearful expressions. The heightened amygdala activity during the processing of happy facial expressions in WS may be a neural marker of anticipatory processes that precede events during which the likelihood of social engagement is high.

Our current study is limited by several constraints defined by our subject groups and experimental design. WS is often associated with general delays in cognitive development. In our ERP study, we took advantage of an IQ-matched (developmentally delayed) control group to demonstrate that the observed responsiveness to social stimuli was not driven by IQ. Due to limited recruitment that yielded poor statistical power, we were not able to use such a control group for our fMRI study. However, we undertook several procedures to investigate whether the observed differences between the WS and TD group in the fMRI study were driven by differences in either IQ or behavioral responses. The findings from these analyses, as well as from the ERP study that included DD controls, indicate that differences in IQ or task performance are not driving the significant between-group fMRI findings in amygdala response to positive and negative social stimuli.

Our data provide evidence that a fundamental characteristic of the WS phenotype consists of aberrant neural reactivity to emotional facial expressions. It is unclear, however, what specific components of facial expressions may correspond with this differential response. Previous studies have reported impaired (Gagliardi et al., 2003; Plesa-Skwerer et al., 2006), as well as preserved (Santos et al., 2008), functioning in face recognition in WS. Additionally, in terms of social cognition, it has been speculated that those with WS exhibit a relatively preserved ability to empathize with others (Mervis and Klein-Tasman, 2000) and that they tend to be relatively socially anxious (Dykens, 2003). Future developments in neuropsychological assessment techniques may strengthen our understanding of the relationship between these psychological constructs and the differences in neural activation reported here.

Advancements in social cognitive neuroscience have demonstrated that brain responses to emotional stimuli vary according to task instructions and states, such as mood (Haas and Canli, 2008). In a previous study, Meyer-Lindenberg et al. (2005) provided evidence of atypical amygdala function during social processing by using a matching task (faces vs shapes). Our study used similar negative (fearful) social stimuli but also included positive (happy) social stimuli and a different set of task instructions (gender discrimination). The convergence between the patterns of results between the two types of tasks strengthens the hypothesis that amygdala abnormality is a primary neural substrate that underlies the social phenotype in WS. Continued research that uses different task instructions and assesses other between-subject characteristics, such as mood and personality, will further advance the existing model of WS and social cognitive brain function.

In summary, this article provides new fMRI and ERP evidence that persons with WS exhibit heightened amygdala and cortical response to happy facial expressions and extends previous research that demonstrated absent or reduced response to fear in WS. This finding provides a neural correlate to previous behavioral findings reporting the tendency to be overly socially appetitive in WS. Future studies will extend our understanding of the genetic influences of social behavior in WS and the interaction between the amygdala and other brain regions associated with social cognition.

References

Aggleton J P (Ed.) (2000) The amygdala: a functional analysis, Ed 2. Oxford: Oxford UP.

Balconi M, Lucciari S (2007) Consciousness and emotional facial expression recognition: subliminal/supraliminal stimulation effect on N200 and P300 ERPs. J Psychophysiol 21:100–108.

Balleine BW, Killcross S (2006) Parallel incentive processing: an integrated view of amygdala function. Trends Neurosci 29:272–279.

Bellugi U, Adolphs R, Cassady C, Chiles M (1999) Towards the neural basis for hypersociability in a genetic syndrome. Neuroreport 10:1653–1657.

Canli T, Sivers H, Whitfield SL, Gotlib IH, Gabrieli JD (2002) Amygdala response to happy faces as a function of extraversion. Science 296:2191.

Costafreda SG, Brammer MJ, David AS, Fu CH (2008) Predictors of amygdala activation during the processing of emotional stimuli: a meta-analysis of 385 PET and fMRI studies. Brain Res Rev 58:57–70.

Derogatis LR (1977) SCL-90: administration, scoring and procedures manual for the revised version and other instruments of the psychopathology rating scale series. Baltimore: John Hopkins University.

Deutsch SI, Rosse RB, Schwartz BL (2007) Williams syndrome: a genetic deletion disorder presenting clues to the biology of sociability and clinical challenges of hypersociability. CNS Spectr 12:903–907.

Diedrich O, Naumann E, Maier S, Becker G (1997) A frontal positive slow wave in the ERP associated with emotional slides. J Psychophysiol 11:71–84.

Dolcos F, Cabeza R (2002) Event-related potentials of emotional memory: encoding pleasant, unpleasant, and neutral pictures. Cogn Affect Behav Neurosci 2:252–263.

Doyle TF, Bellugi U, Korenberg JR, Graham J (2004) "Everybody in the world is my friend" hypersociability in young children with Williams syndrome. Am J Med Genet A 124A:263–273.

Dykens EM (2003) Anxiety, fears, and phobias in persons with Williams syndrome. Dev Neuropsychol 23(1–2):291–316.

Eimer M, Holmes A (2007) Event-related brain potential correlates of emotional face processing. Neuropsychologia 45:15–31.

Feinstein C, Singh S (2007) Social phenotypes in neurogenetic syndromes. Child Adolesc Psychiatr Clin N Am 16:631–647.

Frigerio E, Burt DM, Gagliardi C, Cioffi G, Martelli S, Perrett DI, Borgatti R (2006) Is everybody always my friend? Perception of approachability in Williams syndrome. Neuropsychologia 44:254–259.

Gagliardi C, Frigerio E, Burt DM, Cazzaniga I, Perrett DI, Borgatti R (2003) Facial expression recognition in Williams syndrome. Neuropsychologia 41:733–738.

Haas BW, Canli T (2008) Emotional memory function, personality structure and psychopathology: a neural system approach to the identification of vulnerability markers. Brain Res Rev 58:71–84.

Holland PC, Gallagher M (1999) Amygdala circuitry in attentional and representational processes. Trends Cogn Sci 3:65–73.

Jones W, Bellugi U, Lai Z, Chiles M, Reilly J, Lincoln A, Adolphs R (2000) II. Hypersociability in Williams syndrome. J Cogn Neurosci 12 [Suppl 1]:30–46.

Klein-Tasman BP, Mervis CB (2003) Distinctive personality characteristics of 8-, 9-, and 10-year-olds with Williams syndrome. Dev Neuropsychol 23:269–290.

Koenig S, Mecklinger A (2008) Electrophysiological correlates of encoding and retrieving emotional events. Emotion 8:162–173.

Kuraoka K, Nakamura K (2007) Responses of single neurons in monkey amygdala to facial and vocal emotions. J Neurophysiol 97:1379–1387.

Levitin DJ, Menon V, Schmitt JE, Eliez S, White CD, Glover GH, Kadis J, Korenberg JR, Bellugi U, Reiss AL (2003) Neural correlates of auditory perception in Williams syndrome: an fMRI study. Neuroimage 18:74–82.

Maddux JM, Kerfoot EC, Chatterjee S, Holland PC (2007) Dissociation of attention in learning and action: effects of lesions of the amygdala central nucleus, medial prefrontal cortex, and posterior parietal cortex. Behav Neurosci 121:63–79.

Mervis CB (2003) Williams syndrome: 15 years of psychological research. Dev Neuropsychol 23(1–2):1–12.

Mervis CB, Becerra AM (2007) Language and communicative development in Williams syndrome. Ment Retard Dev Disabil Res Rev 13:3–15.

Mervis CB, Klein-Tasman BP (2000) Williams syndrome: cognition, personality, and adaptive behavior. Ment Retard Dev Disabil Res Rev 6:148–158.

Mervis CB, Morris CA, Klein-Tasman BP, Bertrand J, Kwitny S, Appelbaum LG, Rice CE (2003) Attentional characteristics of infants and toddlers with Williams syndrome during triadic interactions. Dev Neuropsychol 23:243–268.

Meyer-Lindenberg A, Kohn P, Mervis CB, Kippenhan JS, Olsen RK, Morris CA, Berman KF (2004) Neural basis of genetically determined visuospatial construction deficit in Williams syndrome. Neuron 43:623–631.

Meyer-Lindenberg A, Hariri AR, Munoz KE, Mervis CB, Mattay VS, Morris CA, Berman KF (2005) Neural correlates of genetically abnormal social cognition in Williams syndrome. Nat Neurosci 8:991–993.

Mobbs D, Garrett AS, Menon V, Rose FE, Bellugi U, Reiss AL (2004) Anomalous brain activation during face and gaze processing in Williams syndrome. Neurology 62:2070–2076.

Paton JJ, Belova MA, Morrison SE, Salzman CD (2006) The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature 439:865–870.

Plesa-Skwerer D, Faja S, Schofield C, Verbalis A, Tager-Flusberg H (2006) Perceiving facial and vocal expressions of emotion in individuals with Williams syndrome. Am J Ment Retard 111:15–26.

Porter MA, Coltheart M, Langdon R (2007) The neuropsychological basis of hypersociability in Williams and Down syndrome. Neuropsychologia 45:2839–2849.

Riby DM, Hancock PJ (2008) Viewing it differently: social scene perception in Williams syndrome and autism. Neuropsychologia 46:2855–2860.

Sabatinelli D, Bradley MM, Fitzsimmons JR, Lang PJ (2005) Parallel amygdala and inferotemporal activation reflect emotional intensity and fear relevance. Neuroimage 24:1265–1270.

Sabatinelli D, Lang PJ, Keil A, Bradley MM (2007) Emotional perception: correlation of functional MRI and event-related potentials. Cereb Cortex 17:1085–1091.

Santos A, Rondan C, Milne D, Demonet JF, Deruelle C (2008) Social relevance boosts context processing in Williams syndrome. Dev Neuropsychol 33:553–564.

Sato W, Kochiyama T, Yoshikawa S, Matsumura M (2001) Emotional expression boosts early visual processing of the face: ERP recording and its decomposition by independent component analysis. Neuroreport 12:709–714.

Skuse D, Morris J, Lawrence K (2003) The amygdala and development of the social brain. Ann N Y Acad Sci 1008:91–101.

Streit M, Ioannides AA, Liu L, Wölwer W, Dammers J, Gross J, Gaebel W, Müller-Gärtner HW (1999) Neurophysiological correlates of the recognition of facial expressions of emotion as revealed by magnetoencephalography. Brain Res Cogn Brain Res 7:481–491.

Zitzer-Comfort C, Doyle T, Masataka N, Korenberg J, Bellugi U (2007) Nature and nurture: Williams syndrome across cultures. Dev Sci 10:755–762.

15. D. G. Hangauer, A. F. Monzingo, B. W. Matthews, *Biochemistry* **23**, 5730 (1984).
16. O. Dideberg *et al.*, *Nature* **299**, 469 (1982); J. M. Ghuysen, J. Lamotte-Brasseur, B. Joris, G. D. Shockman, *FEBS Lett.* **342**, 23 (1994).
17. W. Stocker and W. Bode, *Curr. Opin. Struct. Biol.* **5**, 383 (1995).
18. K. L. Constantine *et al.*, *J. Mol. Biol.* **223**, 281 (1992); A. R. Pickford, J. R. Potts, J. R. Bright, I. Phan, I. D. Campbell, *Structure* **5**, 359 (1997).
19. I. E. Collier, P. A. Krasnov, A. Y. Strongin, H. Birkedal-Hansen, G. I. Goldberg, *J. Biol. Chem.* **267**, 6776 (1992).
20. L. Bányai, H. Tordai, L. Patthy, *ibid.* **271**, 12003 (1996).
21. F. Willenbrock *et al.*, *Biochemistry* **32**, 4330 (1993); M. W. Olson, D. C. Gervasi, S. Mobashery, R. Fridman, *J. Biol. Chem.* **272**, 29975 (1997); C. M. Overall *et al.*, *ibid.* **274**, 4421 (1999).
22. J. Hodgson, *Biotechnology* **13**, 554 (1995); A. E. Yu, R. E. Hewitt, E. W. Connor, W. G. Stetler-Stevenson, *Drugs Aging* **11**, 229 (1997); S. A. Watson and G. Tierney, *Biodrugs* **9**, 325 (1998).
23. XDS [W. Kabsch, *J. Appl. Crystallogr.* **21**, 916 (1988)]; CCP4 programs [CCP4, Collaborative Computational Project No. 4, Daresbury, UK, *Acta Crystallogr.* D **50**, 760 (1994)]; SFCHECK [A. A. Vagin, J. Richelle, S. J. Wodak, *ibid.* **55**, 191 (1999)]; AMORE [J. Navaza, *Acta Crystallogr.* A **50**, 157 (1994)]; O [T. A. Jones, J.-Y. Zou, S. W. Cowan, M. Kjeldgaard, *ibid.* **47**, 110 (1991)]; and X-PLOR [A. T. Brünger, *X-PLOR, Version 3.1: A System for X-Ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992)].
24. Figures 1, 2A, 3, A and B, and 4A were made with MOLSCRIPT [P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991)] and RASTER3D [E. A. Merrit and M. E. P. Murphy, *Acta Crystallogr.* D **50**, 869 (1994)]. Figures 3, C and D, and 4B were made with GRASP [A. Nichols, K. A. Sharp, B. Honig, *Proteins* **11**, 281 (1991)].
25. Supported by grants from the Swedish Cancer Foundation, EC project BMH4-CT 96-0012, Novo Nordisk Foundation, and Hedlund's Foundation. We thank Tiina Berg, Ilkka Miinalainen, and Kristian Tryggvason for technical assistance with insect cell cultures. We are also grateful to Richard Kahn and Tatjana Sandalova for assistance with the data collection. Beam time was provided by the ESRF.

21 December 1998; accepted 28 April 1999

Genetics of Mouse Behavior: Interactions with Laboratory Environment

John C. Crabbe,[1]* Douglas Wahlsten,[2] Bruce C. Dudek[3]

Strains of mice that show characteristic patterns of behavior are critical for research in neurobehavioral genetics. Possible confounding influences of the laboratory environment were studied in several inbred strains and one null mutant by simultaneous testing in three laboratories on a battery of six behaviors. Apparatus, test protocols, and many environmental variables were rigorously equated. Strains differed markedly in all behaviors, and despite standardization, there were systematic differences in behavior across labs. For some tests, the magnitude of genetic differences depended upon the specific testing lab. Thus, experiments characterizing mutants may yield results that are idiosyncratic to a particular laboratory.

Targeted and chemically induced mutations in mice are valuable tools in biomedical research, especially in the neurosciences and psychopharmacology. Phenotypic effects of a knockout often depend on the genetic background of the mouse strain carrying the mutation (*1*), but the effects of environmental background are not generally known.

Different laboratories commonly employ their own idiosyncratic versions of behavioral test apparatus and protocols, and any laboratory environment also has many unique features. These variations have sometimes led to discrepancies in the outcomes reported by different labs testing the same genotypes for ostensibly the same behaviors (*2*). Previous studies could not distinguish between interactions arising from variations in the test situation itself and those arising from subtle environmental differences among labs. Usually, such differences are eventually resolved by repetition of tests in multiple labs. However, null mutants and transgenic mice are often scarce and tend to be behaviorally characterized in a single laboratory with a limited array of available tests.

We addressed this problem by testing six mouse behaviors simultaneously in three laboratories (Albany, New York; Edmonton, Alberta, Canada; and Portland, Oregon) using exactly the same inbred strains and one null mutant strain (*3*). We went to extraordinary lengths to equate test apparatus, testing protocols, and all possible features of animal husbandry (*4*). One potentially important feature was varied systematically. Because many believe that mice tested after shipping from a supplier behave differently from those reared in-house, we compared mice either shipped or bred locally at the same age (77 days) starting at the same time (0830 to 0900 hours local time on 20 April 1998) in all three labs (*5*). Each mouse was given the same order of tests [Day 1: locomotor activity in an open field; Day 2: an anxiety test, exploration of two enclosed and two open arms of an elevated plus maze; Day 3: walking and balancing on a rotating rod; Day 4: learning to swim to a visible platform; Day 5: locomotor activation after cocaine injection; Days 6 to 11: preference for drinking ethanol versus tap water (*6*)].

Despite our efforts to equate laboratory environments, significant and, in some cases, large effects of site were found for nearly all variables (Table 1). Furthermore, the pattern of strain differences varied substantially among the sites for several tests. Sex differ-

Table 1. Statistical significance and effect sizes for selected variables in the multisite trial. Color of cell depicts Type I error probability or significance of main effects and two-way interactions from 8 × 2 × 3 × 2 analyses of variance: blue, $P < 0.00001$; purple, $P < 0.001$; gold, $P < 0.01$; dashes with no shading, $P > 0.01$. Cell entries are effect sizes, expressed as partial omega squared, the proportion of variance accounted for by the factor or interaction if only that factor were in the experimental design (range = 0 to 1.0). Multiple R^2 (unbiased estimate) gives the proportion of the variance accounted for by all factors. For the water escape task, results are based on only seven strains because most A/J mice never escaped because of wall-hugging. We recognize that the issue of appropriate alpha level correction for multiple comparisons is contentious. Details of the statistical analyses are available on the Web site (*4*), including a discussion of our rationale for presenting uncorrected values in this table.

Task	Measure	Eight Genotypes	Three Sites	Two Sexes	Local vs Shipped	Genotype x Site	Genotype x Sex	Genotype x Ship	Multiple R^2
Open field	Distance in 15 min	.600	.157	----	----	.059	.045	----	.604
Open field	# vertical movements	.788	.281	.039	----	----	----	----	.772
Cocaine	Difference from Day 1	.338	.053	----	----	.086	----	----	.342
Plus maze	Total arm entries	.385	.327	----	----	.210	----	----	.660
Plus maze	Time in open arms	.082	.212	----	----	.066	----	----	.266
Water maze	Mean escape latency	.221	----	----	.026	----	----	----	.177
Alcohol preference	Alcohol consumed (g/kg)	.483	----	.043	----	----	----	----	.451
Body size	Weight (g)	.408	.204	.637	----	.071	.070	----	.698

[1]Portland Alcohol Research Center, Department of Veterans Affairs Medical Center and Department of Behavioral Neuroscience, Oregon Health Sciences University, Portland, OR 97201, USA. [2]Department of Psychology, University of Alberta, Edmonton, Alberta, Canada T6G 2E9. [3]Department of Psychology, State University of New York at Albany, Albany, NY 12222, USA.

*To whom correspondence should be addressed. E-mail: crabbe@ohsu.edu

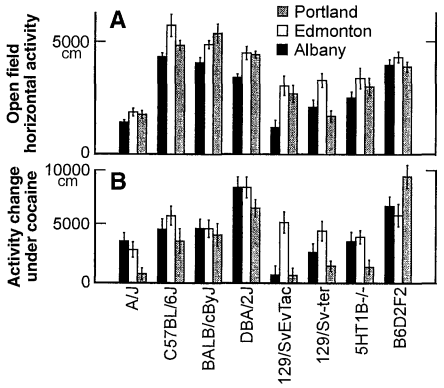


Fig. 1. Group means (±SEM for $n = 16$ mice) for activity in a 40 cm by 40 cm open field for eight strains tested at the same time of day in identical apparatus in three laboratories. (**A**) Horizontal distance (centimeters) traveled in 15 min on the first test on Day 1. (**B**) Cocaine-induced activation, expressed as the difference between horizontal activity (centimeters in 15 min) after cocaine (20 mg/kg) on Day 5 minus the score on Day 1.

ences were only occasionally detected, and, much to our surprise, there were almost no effects of shipping animals before testing. Large genetic effects on all behaviors were confirmed, which is not surprising because we chose strains known to differ markedly on these tasks.

Results for locomotor activity and the effect of a subsequent cocaine injection on locomotion are shown in Fig. 1. Expected strain differences in undrugged activity were found: A/J mice were relatively inactive at all three sites, whereas C57BL/6J mice were very active. An effect of laboratory was also found: mice tested in Edmonton were, on average, more active than those tested in Albany or Portland. In addition, the pattern of genetic differences depended on site. For example, 129/SvEvTac mice tested in Albany were very inactive compared to their counterparts in other labs. Similar results were seen for sensitivity to cocaine stimulation. For example, B6D2F2 mice were very responsive (and A/J mice quite insensitive) to cocaine in Portland, but not at other sites.

In the elevated plus maze, a very similar pattern was seen: strong effects of genotype, site, and their interaction. This was true both for activity measures and for time spent in open arms, the putative index of anxiety (Fig. 2). For total arm entries, the testing laboratory was particularly important for the 5-HT_{1B} knockout mice versus their wild-type 129/Sv-ter background controls. Knockout mice had greater activity than wild types in Portland and tended to have less activity in Albany, while not differing in Edmonton. Edmonton mice of all strains spent more time in open arms (lower anxiety). Portland mice also spent less time in open arms, but this was especially true for strains A/J, BALB/cByJ, and the B6D2F2 mice.

Although the testing laboratory was an important variable, there was a good deal of consistency to the genetic results as well. For example, comparison of the genotype means (averaged over sites) for the initial 5 min of the activity test on Day 1 with the total arm entry scores from the plus maze yielded a high correlation between strains ($r = 0.91$, $P < 0.002$). This indicates that a strain's characteristic activity in novel apparatus is robust and occurs in different apparatus as well as different labs (*7*).

For some behaviors, laboratory environment was not critical. For example, ethanol drinking scores were closely comparable across all three labs, and genotypes alone accounted for 48% of the variance (Table 1 and Fig. 3). The genetic differences showed the well-known pattern of C57BL/6J mice strongly preferring and DBA/2J mice avoiding ethanol (*8*). Females drank more, as is also well known (*8*), but there were no significant effects of site, shipping, or any other interactions. Unlike the other five tests, ethanol preference testing extended over 6 days in the home cage and involved a bare minimum of handling mice by the experimenter.

For some measures, the difference between 5-HT_{1B} null mutant and wild-type mice depended on the specific laboratory environment. In Edmonton, for example, no difference was observed between +/+ and –/– mice in distance traveled in the activity monitor, whereas there was greater activity in the knockouts at the other two sites, especially Portland ($P = 0.002$). In the elevated plus maze, knockouts were considerably more active than wild types only in Portland (Fig. 2A; $P = 0.02$).

The numbers of mice we tested made formal statistical assessment of reliability infeasible, but it would be important to know whether each laboratory would obtain essentially the same strain-specific results if this experiment were repeated. Because our experiment included an internal replication, we estimated the lower bounds of reliability for each site separately by correlating the mean scores for each strain (collapsed over sex and shipping group) obtained during the two replicates of the experiment. These correlations differed depending upon the behavior, and were consonant with the relative importance of genotype in the overall analysis. For example, for locomotor activity, the correlations were 0.97, 0.74, and 0.87 for the three sites. For open-arm time on the plus maze, possibly the most intrinsically unstable task we employed, the correlations were lower (0.32, 0.52, and 0.26). These can be compared to correlations for body weight, which can serve as a type of control variable not influenced by idiosyncratic dynamics of the test situation (0.83, 0.74, and 0.90). No site had generally higher or lower reliability than the others, and formal analyses of replication in analyses of variance indicated no strong interactions of strain by replication. We conclude that reasonable estimates of strain-specific scores are highly dependent on behavioral endpoint, and that some behaviors are highly stable.

Several sources of these laboratory-specific behavioral differences could be ruled out by the rigor of the experimental design. For example, Edmonton mice might have been

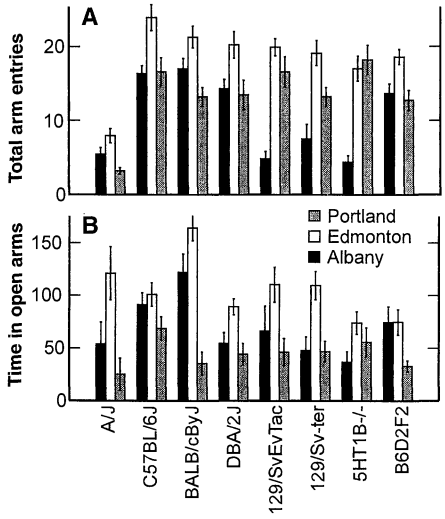


Fig. 2. Group means (±SEM for $n = 16$ mice) for behavior videotaped for 5 min on elevated plus mazes having two open and two enclosed arms. (**A**) Total number of entries into any arm (defined as all four limbs in the arm). (**B**) Time (seconds) spent in the two open arms during the 300-s test. Smaller amounts of time indicate higher levels of anxiety.

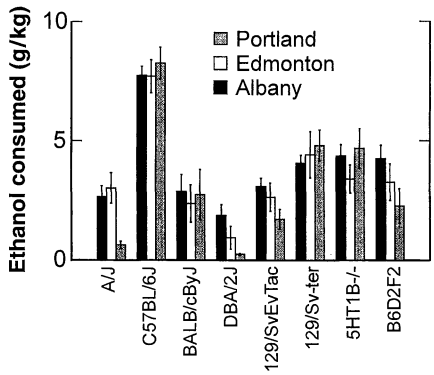


Fig. 3. Mean (±SEM) ethanol consumed per day, expressed as grams per kilogram body weight, over 4 days of an ethanol preference test where each mouse had free access to two drinking bottles, one with local tap water and the other with 6% ethanol in tap water.

more sensitive to cocaine-induced locomotion because the source of cocaine differed from the other two sites (*4*), but this could not explain the relatively marked response of the three 129-derived strains in Edmonton only. However, specific experimenters performing the testing were unique to each laboratory and could have influenced behavior of the mice. The experimenter in Edmonton, for example, was highly allergic to mice and performed all tests while wearing a respirator—a laboratory-specific (and uncontrolled) variable.

Whether animals were bred in each laboratory or shipped as adults 5 weeks before testing had no consistent influence on results in this experiment. Shipped animals took routes of varying duration and difficulty. For example, some Taconic mice were trucked to Albany from nearby Germantown, New York, whereas others spent 2 days in transit during a flight in midwinter to Edmonton. At least in this experiment, allowing animals a lengthy period of acclimation to new quarters was sufficient to overcome any strong effects of putative shipping stress on subsequent behavior.

These results support both optimistic and pessimistic interpretations. Seen optimistically, genotype was highly significant for all behaviors studied, accounting for 30 to 80% of the total variability, and several historically documented strain differences were also seen here. In general, we conclude that very large strain differences are robust and are unlikely to be influenced in a major way by site-specific interactions. However, a more cautious reading suggests that for behaviors with smaller genetic effects (such as those likely to characterize most effects of a gene knockout), there can be important influences of environmental conditions specific to individual laboratories, and specific behavioral effects should not be uncritically attributed to genetic manipulations such as targeted gene deletions.

When studying mutant mice, relatively small genetic effects should first be replicated locally before drawing conclusions (*9*). We further recommend that, if possible, genotypes should be tested in multiple labs and evaluated with multiple tests of a single behavioral domain (such as several tests of anxiety-related behavior) before concluding that a specific gene influences a specific behavioral domain. We also suggest the possibility that laboratory-specific effects on genetic differences will affect phenotypes other than behaviors to an extent similar to that we report.

It is not clear whether standardization of behavioral assays would markedly improve future replicability of results across laboratories. Standardization will be difficult to achieve because most behaviorists seem to have differing opinions about the "best" way to assay a behavioral domain. For example, two of us typically test behavior during the light phase of the animals' cycle, whereas the third typically tests during the dark phase (but switched to the light phase for this study). Which apparatus specifications or test protocol to employ is also a subject of differing opinion. There is a risk of prematurely limiting the "recommended" tests in a domain to those deemed "industry standard," because this may constrain the intrinsic richness of a domain and obscure interesting interactions. On the other hand, increased communication and collaboration between the molecular biologists creating mutations and behavioral scientists interested in the psychological aspects of behavioral testing will benefit both groups.

References and Notes

1. M. Sibilia and E. F. Wagner, *Science* **269**, 234 (1995); R. Gerlai, *Trends Neurosci.* **19**, 177 (1996); M. Nguyen *et al.*, *Nature* **390**, 78 (1997).
2. It has been known for some time that comparisons of multiple genotypes on learning-related tests do not always yield consistent results across laboratories [D. Wahlsten, in *Psychopharmacology of Aversively Motivated Behavior*, H. Anisman and G. Bignami, Eds. (Plenum, New York, 1978), pp. 63–118]. For another example, the Crabbe laboratory has reported that C57BL/6 mice show a small enhancement of locomotor activity after low doses of ethanol, while the Dudek laboratory finds no such stimulant response [J. C. Crabbe *et al.*, *J. Comp. Physiol. Psychol.* **96**, 440 (1982); B. C. Dudek and T. J. Phillips, *Psychopharmacology* **101**, 93 (1990)]. Similar variation has been reported in other measures of activity in various laboratories and apparatus [J. M. LaSalle and D. Wahlsten, in *Techniques for the Genetic Analysis of Brain and Behavior: Focus on the Mouse*, D. Goldowitz, D. Wahlsten, R. E. Wimer, Eds. (Elsevier, Amsterdam, 1992), pp. 391–406].
3. We tested males and females from the inbred strains: A/J, BALB/cByJ, C57BL/6J, DBA/2J, 129/Sv-ter, and 129/SvEvTac; the F_2 hybrid cross of C57BL/6J and DBA/2J (B6D2F2); and the serotonin receptor subtype null mutant, 5-$HT_{1B}^{-/-}$, which is maintained on the 129/Sv-ter background. Mice were obtained from the Jackson Laboratory (Bar Harbor, ME), Taconic Farms (Germantown, NY), or the colonies of R. Hen (Columbia University, New York, NY). Because many targeted deletions are placed on the 129/SvEvTac background, we included this close relative of 129/Sv-ter. The genealogy of many 129 substrains has recently been discussed [E. M. Simpson *et al.*, *Nature Genet.* **16**, 19 (1997); D. W. Threadgill, D. Yee, A. Matin, J. H. Nadeau, T. Magnuson, *Mamm. Genome* **8**, 390 (1997)].
4. Details of procedures and test protocols are given in the Web site for this study (www.albany.edu/psy/obssr). Variables explicitly equated across laboratories included apparatus, exact testing protocols, age of shipped and laboratory-reared mice, method and time of marking before testing, food (Purina 5001; Purina 5000 for breeders), bedding (Bed-o-cob, 1/4 inch; Animal Specialties, Inc., Hubbard, OR), stainless steel cage tops, four to five mice per cage, light/dark cycle, cage changing frequency and specific days, male left in cage after births, culling only of obvious runts, postpartum pregnancy allowed, weaned at 21 days, specific days of body weight recording, and gloved handling without use of forceps. Unmatched variables included local tap water, requirement of filters over cage tops in Portland only, variation of physical arrangement of colonies and testing rooms across sites, different air handling and humidity, and different sources of batches of cocaine and alcohol.
5. All breeding stock was shipped on 2 or 3 December 1997, and mating pairs were set simultaneously on 13 January 1998 in all labs to provide "unshipped" mice for testing. On 15 to 17 March 1998, a second batch of mice from each genotype was shipped to each laboratory. These "shipped" mice, age matched with the unshipped cohort already in place, were allowed to acclimate to the laboratory for 5 weeks before testing commenced. We tested 128 mice in each lab, in two groups of 64 separated by 1 week. With an $n = 4$ mice in each genotype/shipping condition/sex/laboratory condition, we had 16 mice per group for the crucial genotype × laboratory comparisons. This sample size gave us statistical power of 90% to detect modest interactions of genotype × laboratory when Type I error probability was set at 0.01 [J. Cohen, *Statistical Power Analysis* (Erlbaum, Hillsdale, NJ, 1988); D. Wahlsten, *Behav. Brain. Sci.* **13**, 109 (1990)]. For results of analysis of variance, we report only effects significant at $P < 0.01$. The Web site in (*4*) provides detailed protocols used for each test, descriptions of the laboratory conditions rigorously equated across labs, and raw data that may be examined for other interesting patterns.
6. AccuScan Digiscan monitors (AccuScan Instruments, Columbus, OH) were generously loaned to D. Wahlsten by R. H. Kant to match those available in the other two laboratories. AccuScan also provided all sites with rotarod apparatus. Mouse-scaled water mazes and elevated plus mazes were constructed by D. Wahlsten and shipped to the other two labs. On the first test day, each mouse was tested for 15 min in a Digiscan open-field monitor in a dark, sound-attenuated chamber. On Day 2, each mouse was videotaped for 5 min in an elevated plus maze. On Day 3, mice were given 10 trials on a rotarod set to accelerate from 0 to 100 rpm in 75 s. After all mice had been tested on the rotarod, mice were pretrained briefly to escape from the water maze. On Day 4, mice were given eight massed trials of escape learning to a visible platform in the water maze. On Day 5, the activity test was repeated immediately following an ip injection of 20 mg of cocaine per kilogram. After 2 days of rest, mice were individually housed, given only tap water for 2 days, and then tested for 4 days for drinking of 6% ethanol in tap water versus tap water alone.
7. J. Flint *et al.*, *Science* **269**, 1432 (1995); S. R. Mitchell, J. K. Belknap, J. C. Crabbe, unpublished observations.
8. G. E. McClearn and D. A. Rodgers, *Q. J. Stud. Alcohol* **20**, 691 (1959); J. K. Belknap, J. C. Crabbe, E. R. Young, *Psychopharmacology* **112**, 503 (1993); L. A. Rodriguez *et al.*, *Alcohol. Clin. Exp. Res.* **19**, 367 (1995).
9. It was previously reported that 5-HT_{1B} null mutant mice drank much more alcohol than the 129/Sv-ter wild-type strain [J. C. Crabbe *et al.*, *Nature Genet.* **14**, 98 (1996)]. In the experiments here, no site detected this difference (Fig. 3 and Table 1). The original outcome was replicated four times (J. C. Crabbe *et al.*, unpublished data). It is possible that residual polymorphisms for genes segregating in the 129/SvPas substrain that served as the original source of the embryonic stem cell line and in the 129/Sv-ter substrain to which the null mutant was crossed have subsequently been fixed differentially in the 5-HT_{1B} +/+ and –/– strains maintained at Columbia University (*3*). If so, these genes must exert very large epistatic effects on the 1B gene deletion's phenotypic effects on drinking (*1*). Alternatively, some undetected variable (for example, a change in animal care personnel) may have occurred specifically at the Portland site between the original (1995–96) observations and the current experiments.

10. Supported by the Office of Behavioral and Social Sciences Research, NIH, as supplements to grants AA10760 (J.C.C.) and DA10731 (J. Marley and B.C.D., co–principal investigators), and by the Natural Sciences and Engineering Research Council of Canada Grant # 45825 (D.W.), the Department of Veterans Affairs (J.C.C.), and a K02 Award to B.C.D. AA00170. We thank R. H. Kant at AccuScan for the generous loan of equipment and R. Hen for providing the serotonin receptor mutants. We appreciate the comments of C. Cunningham, R. A. Harris, J. Janowsky, and G. Westbrook on a draft of this manuscript. We also thank S. Boehm II, S. Burkhart-Kasch, J. Dorow, S. Doerksen, C. Downing, J. Fogarty, K. Henricks, C. McKinnon, C. Merrill, P. Metten, C. Nolte, T. Phillips, M. Schalomon, J. Schlumbohm, J. Sibert, J. Singh, and C. Wenger for valuable assistance.

1 February 1999; accepted 7 May 1999

nature genetics

ARTICLES

https://doi.org/10.1038/s41588-018-0090-3

Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression

Major depressive disorder (MDD) is a common illness accompanied by considerable morbidity, mortality, costs, and heightened risk of suicide. We conducted a genome-wide association meta-analysis based in 135,458 cases and 344,901 controls and identified 44 independent and significant loci. The genetic findings were associated with clinical features of major depression and implicated brain regions exhibiting anatomical differences in cases. Targets of antidepressant medications and genes involved in gene splicing were enriched for smaller association signal. We found important relationships of genetic risk for major depression with educational attainment, body mass, and schizophrenia: lower educational attainment and higher body mass were putatively causal, whereas major depression and schizophrenia reflected a partly shared biological etiology. All humans carry lesser or greater numbers of genetic risk factors for major depression. These findings help refine the basis of major depression and imply that a continuous measure of risk underlies the clinical phenotype.

MDD is a notably complex and common illness[1]. It is often chronic or recurrent and is thus accompanied by considerable morbidity, disability, excess mortality, substantial costs, and heightened risk of suicide[2–8]. Twin studies attribute approximately 40% of the variation in liability to MDD to additive genetic effects (phenotype heritability, h^2)[9], and h^2 may be greater for recurrent, early-onset, and postpartum MDD[10,11]. Genome-wide association studies (GWAS) of MDD have had notable difficulties in identifying individual associated loci[12]. For example, there were no significant findings in the initial Psychiatric Genomics Consortium (PGC) MDD mega-analysis (9,240 cases)[13] or in the CHARGE meta-analysis of depressive symptoms ($n = 34{,}549$)[14]. More recent studies have proven modestly successful. A study of Han Chinese women (5,303 recurrent MDD cases) identified significant loci[15], a meta-analysis of depressive symptoms (161,460 individuals) identified 2 loci[16], and an analysis of self-reported major depression identified 15 loci (75,607 cases).

There are many reasons why identifying causal loci for MDD has proven difficult[12]. MDD is probably influenced by many genetic loci, each with small effects[17], as are most common diseases[18], including psychiatric disorders[19,20]. Estimates of the proportion of variance attributable to genome-wide SNPs (SNP heritability, h^2_{SNP}) indicate that around one-quarter of the h^2 for MDD is due to common genetic variants[21,22] and demonstrate that a genetic signal is detectable in genome-wide association data, implying that larger sample sizes are needed to detect specific loci given their effect sizes. Such a strategy has been proven in studies of schizophrenia, the flagship adult psychiatric disorder in genomics research. We thus accumulated clinical, population, and volunteer cohorts[23]. This pragmatic approach takes the view that sample size can overcome heterogeneity to identify risk alleles that are robustly associated with major depression. Potential concerns about combining carefully curated research cohorts with volunteer cohorts were considered; multiple lines of evidence suggest that the results are likely to be applicable to clinical MDD. As discussed below, our analyses have neurobiological, clinical, and therapeutic relevance for major depression.

Results

Cohort analyses: phenotype validation. We identified seven cohorts that used a range of methods to ascertain cases with major depression (Table 1 and Supplementary Tables 1–3). The methods used by these cohorts were thoroughly reviewed, drawing on the breadth of expertise in the PGC, and we assessed the comparability of the cohorts using genomic data. We use 'MDD' to refer to directly evaluated subjects meeting standard criteria for major depressive disorder and use 'major depression' where case status was determined using alternative methods as well as to the phenotype from the full meta-analysis.

We evaluated the comparability of the seven cohorts by estimating the common variant genetic correlations (r_g) between them. These analyses supported the comparability of the seven cohorts (Supplementary Table 3), as the weighted mean r_g was 0.76 (s.e. = 0.03). The high genetic correlations between the 23andMe and other cohorts are notable. While there was no statistical evidence of heterogeneity in the r_g estimates for pairs of cohorts ($P = 0.13$), the estimate was statistically different from 1, which may reflect etiological heterogeneity. This estimate can be benchmarked against the slightly larger weighted mean r_g between schizophrenia cohorts of 0.84 (s.e. = 0.05)[21].

Given the positive evidence of the genetic comparability of these cohorts, we completed a genome-wide association meta-analysis of 9.6 million imputed SNPs in 135,458 MDD and major depression cases and 344,901 controls (Fig. 1). There was no evidence of residual population stratification[24] (LD score regression intercept = 1.018, s.e. = 0.009). We estimated h^2_{SNP} to be 8.7% (s.e. = 0.004, liability scale, assuming lifetime risk of 0.15; Supplementary Fig. 1 and Supplementary Table 3b), and note that this is about one-quarter of the h^2 estimated from twin or family studies[9]. This fraction is somewhat lower than that of other complex traits[18] and is plausibly due to etiological heterogeneity (reflecting the mean $r_g < 1$ between cohorts).

To evaluate the impact of combining major depression cohorts that used different ascertainment methods, we undertook a series of genetic risk score (GRS) prediction analyses to demonstrate the validity of our genome-wide association results for clinical MDD (Fig. 2). Notably, the variance explained in out-of-sample prediction increased with the size of the genome-wide association discovery cohort (Fig. 2a), with the GRS from the full discovery

A full list of authors and affiliations appears at the end of the paper.

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

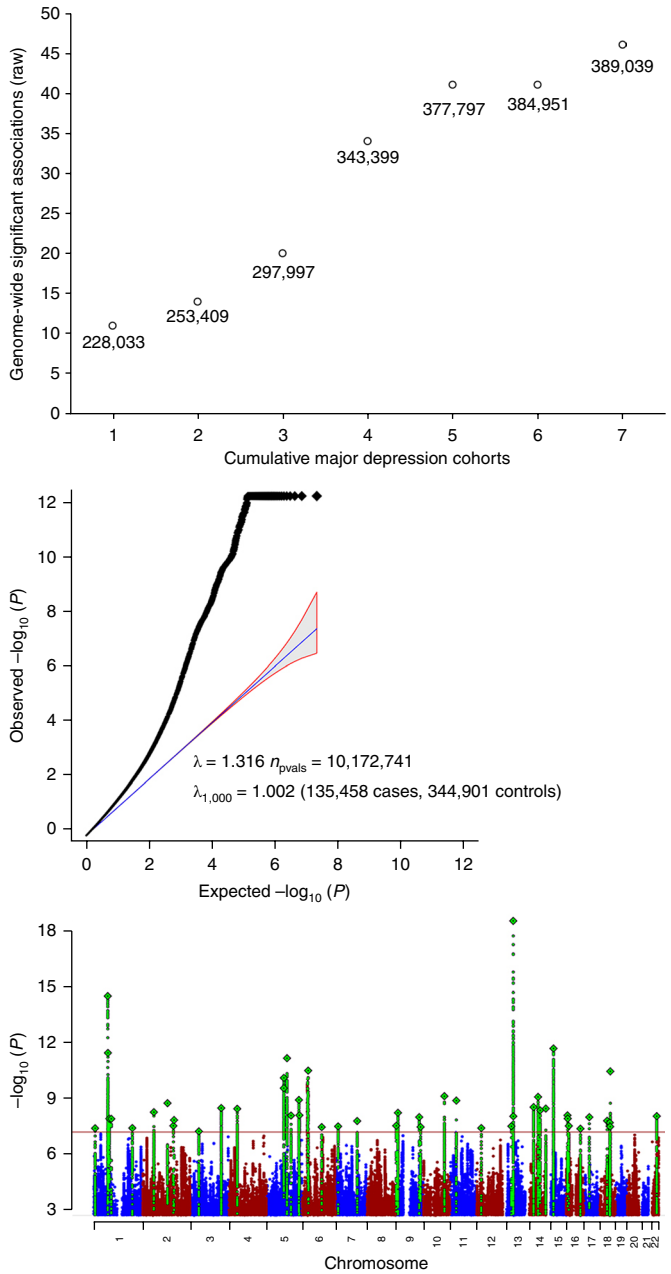


Fig. 1 | Results of genome-wide association meta-analysis of seven cohorts for major depression. a, Relationship between adding cohorts and the number of genome-wide significant genomic regions (before the vetting used to define the final 44 regions). Beginning with the largest cohort (cohort 1 on the *x* axis), we added the next largest cohort (cohort 2) until all cohorts were included (7 cohorts). The number next to each point is the total effective sample size, equivalent to the sample size where the numbers of cases and controls are equal. **b**, Association test quantile–quantile plot showing a marked departure from the null model of no associations (*y* axis truncated at 10^{-12}). **c**, Manhattan plot for association tests from meta-analysis of 135,458 major depression cases and 344,901 controls, with the *x* axis showing genomic position (chromosomes 1–22 plus the X chromosome) and the *y* axis showing statistical significance as $-\log_{10}(P)$ *z* statistics; the threshold for significance accounting for multiple testing is shown by the red horizontal line ($P = 5 \times 10^{-8}$).

sample meta-analysis explaining 1.9% of variance in liability (Fig. 2a, Supplementary Fig. 2, and Supplementary Table 4). For any randomly selected case and control, GRS ranked cases higher than controls with probability of 0.57 and the odds ratio of MDD for those in the tenth versus first GRS decile (OR = 10) was 2.4 (Fig. 2b and Supplementary Table 4). GRS analyses in other disorders (for example, schizophrenia[25]) have shown that the mean GRS increases with clinical severity in cases. We found a significantly higher major depression GRS in those with more severe MDD, as measured in different ways (Fig. 2c). Last, because around one-half of the major depression cases were identified by self-report (i.e., diagnosis or treatment for clinical depression by a medical professional), we further evaluated the comparability of the 23andMe cohort with the other cohorts (full meta-analysis excluding 23andMe, 'FMex23andMe') as detailed in Fig. 2c, Supplementary Table 5, and the Supplementary Note. Taken together, we interpret these results as supporting this meta-analysis of GWAS for these seven cohorts.

Implications for the biology of major depression. Our meta-analysis of seven MDD and major depression cohorts identified 44 independent loci that were statistically significant ($P < 5 \times 10^{-8}$), statistically independent of any other signal[26], and supported by multiple SNPs. This number supports our prediction that genome-wide association discovery in major depression would require about five times more cases than for schizophrenia (lifetime risk ~1% and h^2 ~0.8) to achieve approximately similar power[27]. Of these 44 loci, 30 are new and 14 were significant in a prior study of MDD or depressive symptoms. The overlap of our findings with prior reports was as follows: 1 of 1 with CHARGE depressive symptom[14], 1 of 2 overlap with SSGAC depressive symptom[16], and 12 of 15 overlap with Hyde et al.[28]. There are few trans-ancestry comparisons for major depression, so we contrasted these European results with the Han Chinese CONVERGE study[15] (Supplementary Note). The loci identified in CONVERGE are uncommon in Europeans (rs12415800, 0.45 versus 0.02; rs35936514, 0.28 versus 0.06) and were not significant in our analysis.

Table 2 lists genes in or near the lead SNP in each region, regional plots are in Supplementary Data 1, and Supplementary Tables 6 and 7 provide summaries of available information about the biological functions of the genes in each region. In the Supplementary Note, we review four key genes in more detail: *OLFM4* and *NEGR1* (notable for reported associations with obesity and body mass index[29–34]), *RBFOX1* (notable for independent associations at both the 5′ and 3′ ends, a splicing regulator[35,36], with a functional role that may be consistent with chronic hypothalamic–pituitary–adrenal axis hyperactivation reported in MDD[37]), and *LRFN5* (notable for its role in presynaptic differentiation[38,39] and neuroinflammation[40]).

Gene-wise analyses identified 153 significant genes after controlling for multiple comparisons (Supplementary Table 7). Many of these genes were in the extended major histocompatibility complex (MHC) region (45 of 153), and their interpretation is complicated by high linkage disequilibrium (LD) and gene density. In addition to the genes discussed above, other notable and significant genes outside of the MHC region include multiple potentially 'druggable' targets that suggest connections of the pathophysiology of MDD to neuronal calcium signaling (*CACNA1E* and *CACNA2D1*), dopaminergic neurotransmission (*DRD2*, a principal target of antipsychotics), glutamate neurotransmission (*GRIK5* and *GRM5*), and presynaptic vesicle trafficking (*PCLO*).

Finally, comparison of the major depression loci with 108 loci for schizophrenia[19] identified 6 shared loci. Many SNPs in the extended MHC region are strongly associated with schizophrenia, but implication of the MHC region is new for major depression. Another example is *TCF4* (transcription factor 4), which is strongly associated with schizophrenia but was not previously associated with MDD. TCF4 is essential for normal brain development, and rare mutations in *TCF4* cause Pitt–Hopkins syndrome, which includes autistic features[41]. The GRS calculated from the schizophrenia

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

Table 1 | Brief description of the seven MDD or major depression cohorts used in the meta-analysis

Sample	Country	Case ascertainment	Cases	Controls
PGC29[13,a]	Various	Structured diagnostic interviews[b]	16,823	25,632
deCODE[13]	Iceland	National inpatient electronic records	1,980	9,536
GenScotland[78,79]	UK	Structured diagnostic interview	997	6,358
GERA[80]	USA	Kaiser Permanente Northern California Healthcare electronic medical records (1995–2013)	7,162	38,307
iPSYCH[81]	Denmark	National inpatient electronic records	18,629	17,841
UK Biobank[82] (pilot data release)	UK	From self-reported MDD symptoms or treatment or electronic records[69]	14,260	15,480
23andMe[28] (discovery sample)[c]	USA	Self-reported diagnosis or treatment for clinical depression by a medical professional	75,607	231,747
Total			135,458	344,901

[a]Nineteen samples in addition to the ten samples published in the first PGC-MDD paper[13]. [b]One sample used natural language processing of electronic medical records followed by expert diagnostic review. [c]In Hyde et al.[28], SNPs in 15 genomic regions met genome-wide significance in the combined discovery and replication samples and 11 regions achieved genome-wide significance in the discovery sample made available to the research community and used here. More details are provided in Supplementary Tables 1–3.

genome-wide association results explained 0.8% of the variance in liability of MDD (Fig. 2c).

Implications from integration of functional genomic data. Results from 'omic' studies of functional features of cells and tissues are necessary to understand the biological implications of the results of GWAS for complex disorders[42]. To further elucidate the biological relevance of the major depression findings, we integrated the results with functional genomic data. First, using enrichment analyses, we compared the major depression GWAS findings to bulk tissue mRNA-seq from Genotype-Tissue Expression (GTEx) data[43]. Only brain samples showed significant enrichment (Fig. 3a), and the three tissues with the most significant enrichments were all cortical. Prefrontal cortex and anterior cingulate cortex are important for higher-level executive functions and emotional regulation, which are often impaired in MDD. Both of these regions were implicated in a large meta-analysis of brain magnetic resonance imaging (MRI) findings in adult MDD cases[44]. Second, given the predominance of neurons in the cortex, we confirmed that the major depression genetic findings connect to genes expressed in neurons but not oligodendrocytes or astrocytes (Fig. 3b)[45]. Given the different methods used by the seven MDD/major depression cohorts in this study, demonstration of enrichment of association signals in the brain regions expected to be most relevant to MDD provides independent support for the validity of our approach.

Third, we used partitioned LD score regression[46] to evaluate the enrichment of the major depression genome-wide association findings in over 50 functional genomic annotations (Fig. 3c and Supplementary Table 8). The major finding was the significant enrichment of h^2_{SNP} in genomic regions conserved across 29 Eutherian mammals[47] (20.9-fold enrichment, $P = 1.4 \times 10^{-15}$). This annotation was also the most enriched for schizophrenia[46]. We could not evaluate regions conserved in primates or human 'accelerated' regions, as there were too few for confident evaluation[47]. The other enrichments implied regulatory activity and included open chromatin in human brain and an epigenetic mark of active enhancers (H3K4me1). Notably, exonic regions did not show enrichment, suggesting that, as with schizophrenia[17], genetic variants that change exonic sequences may not have a large role in major depression. We found no evidence that Neanderthal introgressed regions were enriched for major depression genome-wide association findings[48].

Fourth, we applied methods to integrate genome-wide association SNP results with those from gene expression and methylation quantitative trait locus (eQTL and meQTL) studies. SMR[49] analysis identified 13 major depression–associated SNPs with strong evidence that they control local gene expression in one or more tissues and 9 with strong evidence that they control local DNA methylation (Supplementary Table 9 and Supplementary Data 2). A transcriptome-wide association study[50] applied to data from the dorsolateral prefrontal cortex[51] identified 17 genes where major depression–associated SNPs influenced gene expression (Supplementary Table 10). These genes included *OLFM4* (discussed above).

Fifth, we added additional data types to attempt to improve understanding of individual loci. For the intergenic associations, we evaluated total-stranded RNA-seq data from human brain and found no evidence for unannotated transcripts in these regions. A particularly important data type is assessment of DNA–DNA interactions, which can localize a genome-wide association finding to a specific gene that may be nearby or hundreds of kilobases away[52–54]. We integrated the major depression results with 'easy Hi-C' data from brain cortical samples (3 adult, 3 fetal, >1 billion reads each). These data clarified three associations. The statistically independent associations in *NEGR1* (rs1432639, $P = 4.6 \times 10^{-15}$) and over 200 kb away (rs12129573, $P = 4.0 \times 10^{-12}$) both implicate *NEGR1* (Supplementary Fig. 3a), the former likely due to the presence of a reportedly functional copy number polymorphism (see Supplementary Note) and the presence of intergenic loops. The latter association has evidence of DNA looping interactions with *NEGR1*. The association in *SOX5* (rs4074723) and the two statistically independent associations in *RBFOX1* (rs8063603 and rs7198928, $P = 6.9 \times 10^{-9}$ and 1.0×10^{-8}) had only intragenic associations, suggesting that the genetic variation in the regions of the major depression associations act locally and can be assigned to these genes. In contrast, the association in *RERE* (rs159963, $P = 3.2 \times 10^{-8}$) could not be assigned to *RERE* as it may contain super-enhancer elements given its many DNA–DNA interactions with many nearby genes (Supplementary Fig. 3b).

Implications based on the roles of sets of genes. A parsimonious explanation for the presence of many significant associations for a complex trait is that the different associations are part of a higher-order grouping of genes[55]. These could be a biological pathway or a collection of genes with a functional connection. Multiple methods allow evaluation of the connection of major depression genome-wide association results to sets of genes grouped by empirical or predicted function (pathway or gene set analysis).

Full pathway analyses are in Supplementary Table 11, and 19 pathways with false discovery rate q values <0.05 are summarized in Fig. 4. The major groupings of significant pathways were as follows: RBFOX1, RBFOX2, RBFOX3, or CELF4 regulatory networks; genes whose mRNAs are bound by FMRP; synaptic genes; genes

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

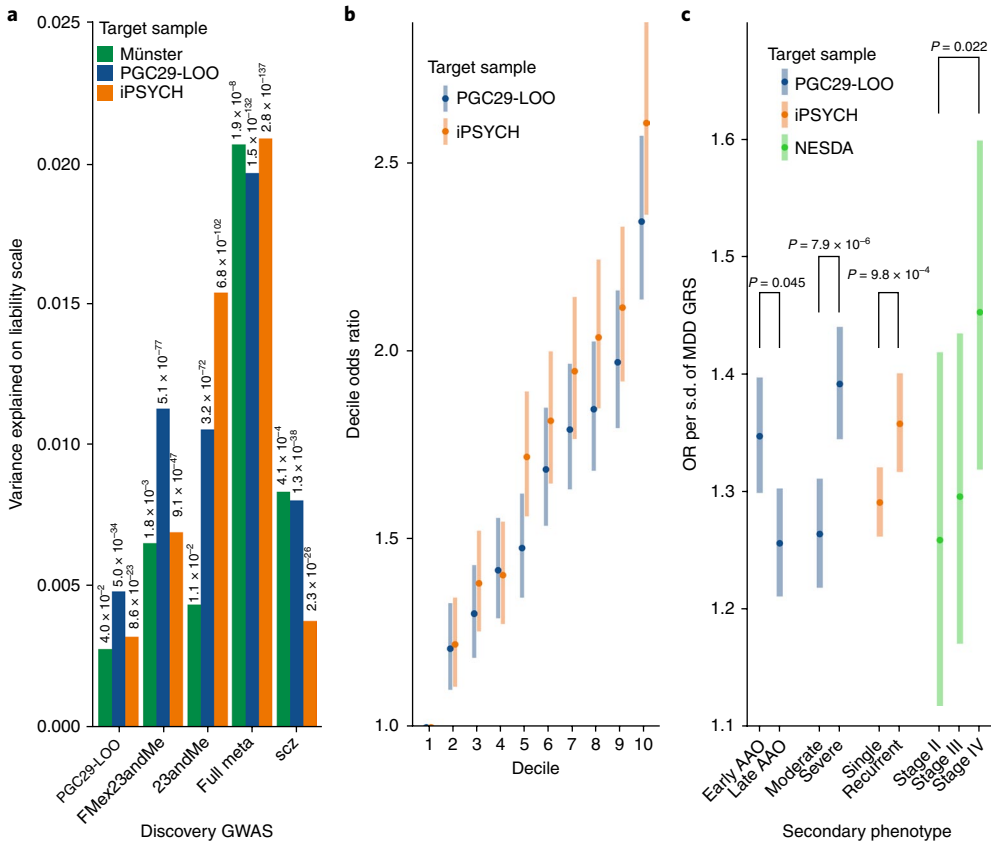


Fig. 2 | Genetic risk score prediction analyses into PGC29 MDD target samples. a, Variance explained (liability scale) based on different discovery samples for three target samples: PGC29 (16,823 cases, 25,632 controls), iPSYCH (a nationally representative sample of 18,629 cases and 17,841 controls), and a clinical cohort from Münster not included in the genome-wide association analysis (845 MDD inpatient cases, 834 controls). For PGC29-LOO, the target sample was one of the PGC29 samples with the discovery sample being the remaining 28 PGC29 samples, hence representing leave-one-out (LOO) analysis. **b**, Odds ratios of major depression per GRS decile relative to the first decile for the iPSYCH and PGC29 target samples. **c**, Odds ratios of major depression in GRS s.d.: PGC29-LOO, 3,950 early-onset versus 3,950 late-onset cases; PGC29-LOO, 4,958 severe versus 3,976 moderate cases defined by count of endorsed MDD symptom criteria; iPSYCH, 5,574 cases of recurrent MDD versus 12,968 single-episode cases; and NESDA from PGC29, severity defined as chronic/unremitting MDD, 610 'stage IV' cases versus 499 'stage II' or 332 first-episode MDD cases[77] . Error bars represent 95% confidence intervals. Logistic regression association test *P* values are shown in the target sample for the GRS generated from SNPs with $P < 0.05$ in the discovery sample. FMex23andMe, full meta-analysis excluding 23andMe; scz, schizophrenia[19].

involved in neuronal morphogenesis; genes involved in neuron projection; genes associated with schizophrenia (at $P < 10^{-4}$)[19]; genes involved in central nervous system (CNS) neuron differentiation; genes encoding voltage-gated calcium channels; genes involved in cytokine and immune response; and genes known to bind to the retinoid X receptor. Several of these pathways are implicated by GWAS of schizophrenia and by rare exonic variation of schizophrenia and autism[56,57] and immediately suggest shared biological mechanisms across these disorders.

A key issue for common variant GWAS is their relevance for pharmacotherapy. We conducted gene set analysis that compared the major depression genome-wide association results to targets of antidepressant medications defined by pharmacological studies[58] and found that 42 sets of genes encoding proteins bound by antidepressant medications were highly enriched for smaller major depression association *P* values than expected by chance (42 drugs, rank enrichment test $P = 8.5 \times 10^{-10}$). This finding connects our major depression genomic findings to MDD therapeutics and suggests the salience of these results for new lead compound discovery for MDD[59].

Implications based on relationships with other traits. Prior epidemiological studies associated MDD with many other diseases and traits. Because of limitations inherent to observational studies, understanding whether a phenotypic correlation is potentially causal or if it results from reverse causation or confounding is generally difficult. Genetic studies now offer complementary strategies to assess whether a phenotypic association between MDD and a risk factor or a comorbidity is mirrored by a nonzero r_g (common variant genetic correlation) and, for some of these, evaluate the potential causality of the association given that exposure to genetic risk factors begins at conception.

We used LD score regression to estimate the r_g of major depression with 221 psychiatric disorders, medical diseases, and human traits[22,60]. Supplementary Table 12 contains the full results, and Table 3 shows the r_g values with false discovery rates <0.01. First, the r_g values were very high between our major depression genome-wide association results and those from two studies of current depressive symptoms. Both correlations were close to 1 (the samples in one report overlapped partially with this meta-analysis[16], but the samples from the other did not[14]).

Second, we found significant positive genetic correlations between major depression and every psychiatric disorder assessed along with smoking initiation. This is a comprehensive and well-powered evaluation of the relationship of MDD with other psychiatric disorders, and these results indicate that the common genetic

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

Table 2 | 44 significantly associated genomic regions in meta-analysis of 135,458 major depression cases and 344,901 controls

Chr.	Region (Mb)	SNP	Location (bp)	*P*	A1/A2	OR (A1)	s.e. (log(OR))	Freq.	Prev.	Gene context
1	8.390–8.895	rs159963	8,504,421	3.2×10^{-8}	A/C	0.97	0.0049	0.56	H,S	[*RERE*]; *SLC45A1*, 100,194
1	72.511–73.059	rs1432639	72,813,218	4.6×10^{-15}	A/C	1.04	0.0050	0.63	H	*NEGR1*, –64,941
1	73.275–74.077	rs12129573	73,768,366	4.0×10^{-12}	A/C	1.04	0.0050	0.37	S	*LINC01360*, –3,486
1	80.785–80.980	rs2389016	80,799,329	1.0×10^{-8}	T/C	1.03	0.0053	0.28	H	
1	90.671–90.966	rs4261101	90,796,053	1.0×10^{-8}	A/G	0.97	0.0050	0.37		
1	197.343–197.864	rs9427672	197,754,741	3.1×10^{-8}	A/G	0.97	0.0058	0.24		*DENND1B*, –10,118
2	57.765–58.485	rs11682175	57,987,593	4.7×10^{-9}	T/C	0.97	0.0048	0.52	H,S	*VRK2*, –147,192
2	156.978–157.464	rs1226412	157,111,313	2.4×10^{-8}	T/C	1.03	0.0059	0.79		[*LINC01876*]; *NR4A2*, 69,630; *GPD2*, –180,651
3	44.222–44.997	chr3_44287760_I	44,287,760	4.6×10^{-8}	I/D	1.03	0.0051	0.34	T	[*TOPAZ1*]; *TCAIM*, –91,850; *ZNF445*, 193,501
3	157.616–158.354	rs7430565	158,107,180	2.9×10^{-9}	A/G	0.97	0.0048	0.58	H	[*RSRC1*]; *LOC100996447*, 155,828; *MLF1*, –181,772
4	41.880–42.189	rs34215985	42,047,778	3.1×10^{-9}	C/G	0.96	0.0063	0.24		[*SLC30A9*]; *LINC00682*, –163,150; *DCAF4L1*, 59,294
5	87.443–88.244	chr5_87992715_I	87,992,715	7.9×10^{-11}	I/D	0.97	0.0050	0.58	H	*LINC00461*, –12,095; *MEF2C*, 21,342
5	103.672–104.092	chr5_103942055_D	103,942,055	7.5×10^{-12}	I/D	1.03	0.0048	0.48	C	
5	124.204–124.328	rs116755193	124,251,883	7.0×10^{-9}	T/C	0.97	0.0050	0.38		*LOC101927421*, –120,640
5	164.440–164.789	rs11135349	164,523,472	1.1×10^{-9}	A/C	0.97	0.0048	0.48	H	
5	166.977–167.056	rs4869056	166,992,078	6.8×10^{-9}	A/G	0.97	0.0050	0.63		[*TENM2*]
6	27.738–32.848	rs115507122	30,737,591	3.3×10^{-11}	C/G	0.96	0.0063	0.18	S	Extended MHC
6	99.335–99.662	rs9402472	99,566,521	2.8×10^{-8}	A/G	1.03	0.0059	0.24		*FBXL4*, –170,672; *C6orf168*, 154,271
7	12.154–12.381	rs10950398	12,264,871	2.6×10^{-8}	A/G	1.03	0.0049	0.41		[*TMEM106B*]; *VWDE*, 105,637
7	108.925–109.230	rs12666117	109,105,611	1.4×10^{-8}	A/G	1.03	0.0048	0.47		
9	2.919–3.009	rs1354115	2,983,774	2.4×10^{-8}	A/C	1.03	0.0049	0.62	H	*PUM3*, –139,644; *LINC01231*, –197,814
9	11.067–11.847	rs10959913	11,544,964	5.1×10^{-9}	T/G	1.03	0.0057	0.76		
9	119.675–119.767	rs7856424	119,733,595	8.5×10^{-9}	T/C	0.97	0.0053	0.29		[*ASTN2*]
9	126.292–126.735	rs7029033	126,682,068	2.7×10^{-8}	T/C	1.05	0.0093	0.07		[*DENND1A*]; *LHX2*, –91,820
10	106.397–106.904	rs61867293	106,563,924	7.0×10^{-10}	T/C	0.96	0.0061	0.20	H	[*SORCS3*]
11	31.121–31.859	rs1806153	31,850,105	1.2×10^{-9}	T/G	1.04	0.0059	0.22		[*DKFZp686K1684*]; [*PAUPAR*]; *ELP4*, 44,032; *PAX6*, –10,596
12	23.924–24.052	rs4074723	23,947,737	3.1×10^{-8}	A/C	0.97	0.0049	0.41		[*SOX5*]
13	44.237–44.545	rs4143229	44,327,799	2.5×10^{-8}	A/C	0.95	0.0091	0.92		[*ENOX1*]; *LACC1*, –125,620; *CCDC122*, 82,689
13	53.605–54.057	rs12552	53,625,781	6.1×10^{-19}	A/G	1.04	0.0048	0.44	H	[*OLFM4*]; *LINC01065*, 80,099
14	41.941–42.320	rs4904738	42,179,732	2.6×10^{-9}	T/C	0.97	0.0049	0.57		[*LRFN5*]
14	64.613–64.878	rs915057	64,686,207	7.6×10^{-10}	A/G	0.97	0.0049	0.42		[*SYNE2*]; *MIR548H1*, –124,364; *ESR2*, 7,222
14	75.063–75.398	chr14_75356855_I	75,356,855	3.8×10^{-9}	D/I	1.03	0.0049	0.49		[*DLST*]; *PROX2*, –26,318; *RPS6KL1*, 13,801
14	103.828–104.174	rs10149470	104,017,953	3.1×10^{-9}	A/G	0.97	0.0049	0.49	S	*BAG5*, 4,927; *APOPT1*, –11,340
15	37.562–37.929	rs8025231	37,648,402	2.4×10^{-12}	A/C	0.97	0.0048	0.57	H	

Continued

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

Table 2 | 44 significantly associated genomic regions in meta-analysis of 135,458 major depression cases and 344,901 controls (continued)

Chr.	Region (Mb)	SNP	Location (bp)	*P*	A1/A2	OR (A1)	s.e. (log(OR))	Freq.	Prev.	Gene context
16	6.288–6.347	rs8063603	6,310,645	6.9×10^{-9}	A/G	0.97	0.0053	0.65		[*RBFOX1*]
16	7.642–7.676	rs7198928	7,666,402	1.0×10^{-8}	T/C	1.03	0.0050	0.62		[*RBFOX1*]
16	13.022–13.119	rs7200826	13,066,833	2.4×10^{-8}	T/C	1.03	0.0055	0.25		[*SHISA9*]; *CPPED1*, –169,089
16	71.631–72.849	rs11643192	72,214,276	3.4×10^{-8}	A/C	1.03	0.0049	0.41		*PMFBP1*, –7,927; *DHX38*, 67,465
17	27.345–28.419	rs17727765	27,576,962	8.5×10^{-9}	T/C	0.95	0.0088	0.92		[*CRYBA1*]; *MYO18A*, –69,555; *NUFIP2*, 5,891
18	36.588–36.976	rs62099069	36,883,737	1.3×10^{-8}	A/T	0.97	0.0049	0.42		[*MIR924HG*]
18	50.358–50.958	rs11663393	50,614,732	1.6×10^{-8}	A/G	1.03	0.0049	0.45	O	[*DCC*]; *MIR4528*, –148,738
18	51.973–52.552	rs1833288	52,517,906	2.6×10^{-8}	A/G	1.03	0.0054	0.72		[*RAB27B*]; *CCDC68*, 50,833
18	52.860–53.268	rs12958048	53,101,598	3.6×10^{-11}	A/G	1.03	0.0051	0.33	S	[*TCF4*]; *MIR4529*, –44,853
22	40.818–42.216	rs5758265	41,617,897	7.6×10^{-9}	A/G	1.03	0.0054	0.28	H,S	[*L3MBTL2*]; *EP300-AS1*, –24,392; *CHADL*, 7,616

Chr. (chromosome) and region (boundaries in Mb; hg19) are shown, defined by the locations of SNPs with $P < 1 \times 10^{-5}$ and LD $r^2 > 0.1$ with the most strongly associated SNP (logistic regression; the lowest *P* value in the region listed is not corrected for multiple testing), whose location is given in base pairs. In three regions, a second SNP fulfilled the filtering criteria and was followed up with conditional analyses: chr. 1, conditional analysis selected only rs1432639 as significant, with $P = 2.0 \times 10^{-4}$ for rs12134600 after fitting rs1432639; chr. 5, conditional analysis showed two independent associations selecting rs247910 and rs10514301 as the most strongly associated SNPs; chr. 10, conditional analysis selected only rs61867293 with $P = 8.6 \times 10^{-5}$ for rs1021363 after conditioning on rs61867293. For each of the 47 SNPs, there was at least one additional genome-wide significant SNP in the cluster of surrounding SNPs with low *P* values. Chromosome X was analyzed but had no findings that met genome-wide significance. Entries in the "Prev." column indicate which of four previous studies identified genome-wide significant associations in a region: H, Hyde et al.[28], 23andMe genome-wide association of self-reported clinical depression (the discovery sample overlaps with this paper); O, Okbay et al.[16], meta-analysis of genome-wide association of MDD, depressive symptoms, psychological well-being, and neuroticism (includes many PGC29 samples); S, PGC report on 108 schizophrenia-associated loci[19]; C, CHARGE Consortium meta-analysis of depressive symptoms[14]. The "Gene context" column shows the distances between the peak SNP and the closest genes; brackets indicate that the peak SNP was within that gene. The closest genes upstream (taking strand into account; a negative number indicates distance in base pairs between the peak SNP and the nearest gene boundary) and downstream (positive distance in base pairs) are also shown if there was a flanking gene within 200 kb. The name of the closest gene is in brackets. Note that it is generally not known whether associated SNPs have biological effects on these or other more distant genes. A1/A2, the two alleles (or insertion–deletion) at each SNP; A1 was tested for association and its OR (odds ratio) and s.e. are shown; Freq., the frequency of A1 in controls across all cohorts.

variants that predispose to MDD overlap substantially with those for adult- and childhood-onset psychiatric disorders, although they remain substantially distinct as well.

Third, the common variant genetic architecture of major depression was positively correlated with multiple measures of sleep quality (daytime sleepiness, insomnia, and tiredness). The first two of these correlations used UK Biobank data with people endorsing major depression or other major psychiatric disorders, shift workers, and those taking hypnotics excluded. This pattern of correlations combined with the importance of sleep and fatigue in major depression (two criteria for MDD) suggests a close and potentially profound mechanistic relationship. Major depression also had a strong genetic correlation with neuroticism (a personality dimension assessing the degree of emotional instability); this is consistent with the literature showing a close interconnection of MDD and this personality trait. The strong negative r_g with subjective well-being underscores the capacity of major depression to impact human health.

Finally, major depression had significant negative genetic correlations with data from two studies of educational attainment, which while often considered at the genetic level as proxy measures of intelligence also likely includes more complex personality constructs. With this in mind, it is relevant to note that the r_g between major depression and IQ[61] was not significantly different from zero, despite the r_g between years of education and IQ being 0.7, implying complex relationships between these traits worthy of future investigation. We also found significant positive correlations with multiple measures of adiposity, relationship to female reproductive behavior (decreased age at menarche, age at first birth, and increased number of children), and positive correlations with coronary artery disease and lung cancer.

We used bidirectional Mendelian randomization (MR) to investigate the relationships between four traits genetically correlated with major depression: years of education (EDY)[62], body mass index (BMI)[29], coronary artery disease (CAD)[63], and schizophrenia[19]. These traits were selected because all of the following were true: they were phenotypically associated with MDD, had significant r_g with MDD, and had >30 independent genome-wide significant associations from large GWAS. We report GSMR[64] results but obtained qualitatively similar results with other MR methods (Supplementary Fig. 4 and Supplementary Table 13). MR analyses provided evidence for a 1.12-fold increase in major depression per s.d. of BMI ($P_{GSMR} = 1.2 \times 10^{-7}$) and a 0.84-fold decrease in major depression per s.d. of EDY ($P_{GSMR} = 2.3 \times 10^{-6}$). There was no evidence of reverse causality of major depression for BMI ($P_{GSMR} = 0.53$) or EDY ($P_{GSMR} = 0.11$). For BMI, there was some evidence of pleiotropy, as six BMI-associated SNPs were excluded by the HEIDI outlier test including SNPs near *OLFM4* and *NEGR1*. Thus, these results are consistent with EDY and BMI as either causal risk factors or correlated with causal risk factors for major depression. These results provide hypotheses for future research to understand these potentially directional relationships.

For CAD, the MR analyses were not significant when considering major depression as an outcome ($P_{GSMR} = 0.30$) or as an exposure ($P_{GSMR} = 0.12$); however, the high standard error of the estimates using MDD SNP instruments implies that this analysis should be revisited when more major genome-wide significant SNP instruments for depression become available from future GWAS.

We used MR to investigate the relationship between major depression and schizophrenia. Although major depression had positive r_g with many psychiatric disorders, only schizophrenia had

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

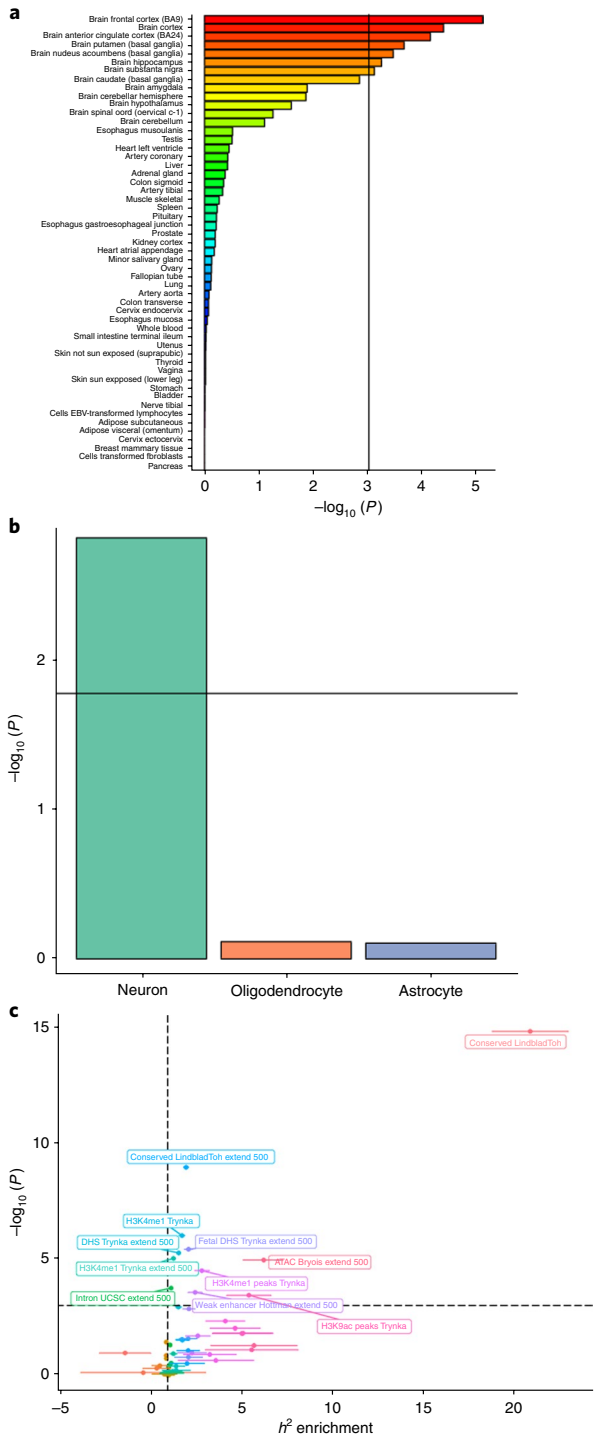


Fig. 3 | Analyses exploring enrichment of major depression association results based on different SNP annotations. a, Enrichment in bulk tissue mRNA-seq data from GTEx shown as *t* statistics; sample sizes in the GTEx database range from $n = 75$–564. The threshold for significance accounting for multiple testing is shown by the vertical line. **b**, Major depression results and enrichment in three major brain cell types shown as *t* statistics; the threshold for significance accounting for multiple testing is shown by the horizontal line. Sample sizes vary as these data were aggregated from many different sources. **c**, Partitioned LDSC to evaluate enrichment of the major depression genome-wide association findings in over 50 functional genomic annotations (Supplementary Table 8) shown as enrichment statistics; the threshold for significance accounting for multiple testing is shown by the horizontal dashed line. Sample sizes vary as these data were aggregated from many different sources.

Table 3 | LDSC genetic correlations of MDD with other disorders, diseases, and human traits

Trait	r_g	s.e.	FDR	h^2_{SNP}	Ref.
Depressive symptoms, CHARGE	0.91	0.123	3.2×10^{-12}	0.04	14
Depressive symptoms, SSGAC	0.98	0.034	1.3×10^{-176}	0.05	16
ADHD (iPSYCH-PGC)	0.42	0.033	6.1×10^{-36}	0.24	83
Anorexia nervosa	0.13	0.028	7.1×10^{-5}	0.55	84
Anxiety disorders	0.80	0.140	2.0×10^{-7}	0.06	85
Autism spectrum disorders (iPSYCH-PGC)	0.44	0.039	8.4×10^{-28}	0.20	86
Bipolar disorder	0.32	0.034	3.3×10^{-19}	0.43	20
Schizophrenia	0.34	0.025	7.7×10^{-40}	0.46	19
Smoking, ever versus never	0.29	0.038	7.0×10^{-13}	0.08	87
Daytime sleepiness[a]	0.19	0.048	5.7×10^{-4}	0.05	–
Insomnia[a]	0.38	0.038	4.0×10^{-22}	0.13	–
Tiredness	0.67	0.037	6.2×10^{-72}	0.07	88
Subjective well-being	−0.65	0.035	7.5×10^{-76}	0.03	16
Neuroticism	0.70	0.031	2.5×10^{-107}	0.09	16
College completion	−0.17	0.034	6.7×10^{-6}	0.08	89
Years of education	−0.13	0.021	1.6×10^{-8}	0.13	62
Body fat	0.15	0.038	6.5×10^{-4}	0.11	90
Body mass index	0.09	0.026	3.6×10^{-3}	0.19	32
Obesity class 1	0.11	0.029	1.6×10^{-3}	0.22	30
Obesity class 2	0.12	0.033	3.0×10^{-3}	0.18	30
Obesity class 3	0.20	0.053	1.6×10^{-3}	0.12	30
Overweight	0.13	0.030	1.4×10^{-4}	0.11	30
Waist circumference	0.11	0.024	8.2×10^{-5}	0.12	91
Waist-to-hip ratio	0.12	0.030	2.9×10^{-4}	0.11	91
Triglycerides	0.14	0.028	1.0×10^{-5}	0.17	92
Age at menarche	−0.14	0.023	6.3×10^{-8}	0.20	93
Age of first birth	−0.29	0.029	6.1×10^{-22}	0.06	94
Father's age at death	−0.28	0.058	3.0×10^{-5}	0.04	95
Number of children ever born	0.13	0.036	2.4×10^{-3}	0.03	94
Coronary artery disease	0.12	0.027	8.2×10^{-5}	0.08	63
Squamous cell lung cancer	0.26	0.075	3.6×10^{-3}	0.04	96

All genetic correlations (r_g) estimated using bivariate LDSC applied to major depression genome-wide association results are in Supplementary Table 12. Shown above are the r_g values of major depression with false discovery rate (FDR) < 0.01 (FDR estimated for 221 genetic correlations; H_0: $r_g = 0$). Thematically related traits are indicated by bold font. iPSYCH is a nationally representative cohort based on blood spots collected at birth. Within iPSYCH, r_g with ADHD was 0.58 (s.e. = 0.050) and with ASD was 0.51 (s.e. = 0.07); these values are larger than those listed above and are inconsistent with artifactual correlations. h^2_{SNP} is shown to aid interpretation as high r_g in the context of high h^2_{SNP} is more noteworthy than when h^2_{SNP} is low. [a]Self-reported daytime sleepiness and insomnia from UK Biobank excluding subjects with major depression, other psychiatric disorders (bipolar disorder, schizophrenia, autism, intellectual disability), shift workers, and those taking hypnotics.

sufficient associations for MR analyses. We found significant bidirectional correlations in SNP effect sizes for schizophrenia loci in major depression ($P_{GSMR} = 1.1 \times 10^{-40}$) and for major depression loci in schizophrenia ($P_{GSMR} = 1.5 \times 10^{-11}$). These results suggest that the major depression–schizophrenia r_g of 0.34 is consistent with

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

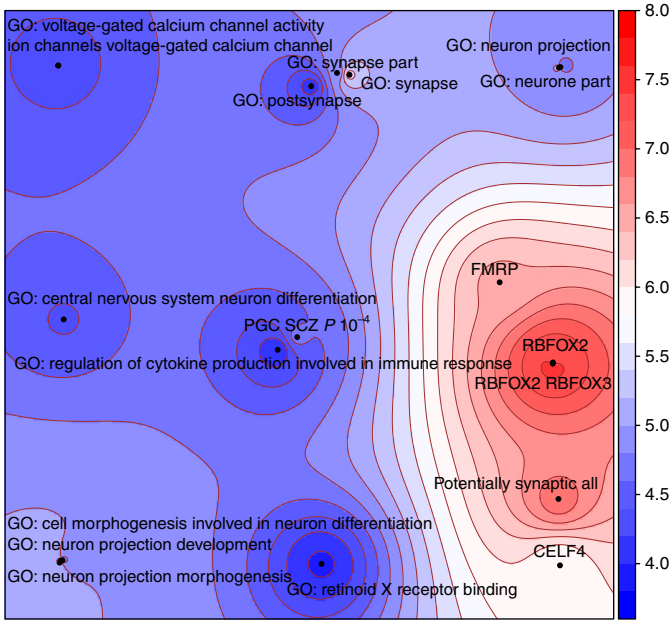


Fig. 4 | Generative topographic mapping of the 19 significant pathway results. The average position of each pathway on the map is represented by a point. The map is colored by the $-\log_{10}$ *P* value obtained using MAGMA. The *x* and *y* coordinates were obtained from a kernel generative topographic mapping algorithm (GTM) that reduces high-dimensional gene sets to a 2D scatterplot by accounting for gene overlap between gene sets. Each point represents a gene set. Nearby points are more similar in gene overlap than more distant points. The color surrounding each point (gene set) corresponds to significance according to the scale on the right. The significant pathways (Supplementary Table 11) fall into nine main clusters as described in the text.

partially shared biological pathways being causal for both disorders. Although it is plausible that diagnostic misclassification/ambiguity (for example, misdiagnosis of MDD as schizoaffective disorder) could contaminate these analyses, levels of misclassification would need to be implausibly high (30% unidirectional, 15% bidirectional) to result in an r_g of ~0.3[65].

All MR analyses were repeated after excluding the 23 and Me cohort, and the pattern of the results was the same (Supplementary Table 13).

Discussion

The nature of severe depression has been discussed for millennia[66]. This genome-wide association meta-analysis is among the largest ever conducted in psychiatric genetics and provides a body of results that helps refine the fundamental basis of major depression.

In conducting this meta-analysis of major depression, we employed a pragmatic approach by including cohorts that met empirical criteria for sufficient genetic and phenotypic similarity. Our approach was cautious, clinically informed, guided by empirical data, and selective (for example, we did not include cohorts with bipolar disorder (which requires MDD), depressive symptoms, neuroticism, or well-being). Approximately 44% of all major depression cases were assessed using traditional methods (PGC29, GenScot), treatment registers (iPSYCH, GERA; such approaches have been extensively used to elucidate the epidemiology of major depression), or a combination of methods (deCODE, UK Biobank), whereas ~56% of cases were from 23andMe (via self-report)[28]. Multiple lines of genetic evidence supported conducting meta-analysis of these seven cohorts (for example, out-of-sample prediction, sign tests, and genetic correlations).

However, our approach may be controversial to some readers given the unconventional reliance on self-report of major depression. We would reframe the issue: we hypothesize that brief methods of assessing major depression are informative for the genetics of MDD. We present a body of results that is consistent with this hypothesis. Even if unconventional, our hypothesis is testable and falsifiable, and we invite and welcome empirical studies to further support or refute this hypothesis.

Our results lead us to draw some broad conclusions. First, major depression is a brain disorder. Although this is not unexpected, some past models of MDD have had little or no place for heredity or biology. The genetic results best match gene expression patterns in the prefrontal and anterior cingulate cortex, anatomical regions that show differences between MDD cases and controls. The genetic findings implicated neurons (not microglia or astrocytes), and we anticipate more detailed cellular localization when sufficient single-cell and single-nucleus RNA-seq datasets become available[67].

Second, the genetic associations for major depression (as with schizophrenia)[46] tend to occur in genomic regions conserved across a range of placental mammals. Conservation suggests important functional roles. Notably, our analyses did not implicate exons or coding regions.

Third, the results also implicated developmental gene regulatory processes. For instance, the genetic findings pointed at the splicing regulator *RBFOX1* (the presence of two independent genetic associations in *RBFOX1* strongly suggests that it is the relevant gene). Gene set analyses implicated genes containing binding sites to the protein product of *RBFOX1*, and this gene set is also significantly enriched for rare exonic variation in autism and schizophrenia[56,57]. These analyses highlight the potential importance of splicing to generate alternative isoforms; risk for major depression may be mediated not by changes in isolated amino acids but rather by changes in the proportions of isoforms coming from a gene, given that isoforms often have markedly different biological functions[68,69]. These convergent results provide possible clues to a biological mechanism common to multiple severe psychiatric disorders that merits future research.

Fourth, in the most extensive analysis of the genetic 'connections' of major depression with a wide range of disorders, diseases, and human traits, we found significant positive genetic correlations with measures of body mass and negative genetic correlations with years of education, while showing no evidence of genetic correlation with IQ. MR analysis results are consistent with both BMI and years of education being causal, or correlated with causal, risk factors for major depression, and our results provide hypotheses and motivation for more detailed prospective studies, as currently available data may not provide insight about the fundamental driver or drivers of causality. The underlying mechanisms are likely more complex, as it is difficult to envision how genetic variation in educational attainment or body mass alters risk for MDD without invoking an additional mechanistic component. While the significant MR analyses need further investigations to fully understand, the negative MR results provide important evidence that there is not a direct causal relationship between MDD and subsequent changes in body mass or education years. If such associations are observed in epidemiological or clinical samples, then other factors must drive the association.

Fifth, we found significant positive correlations of major depression with all psychiatric disorders that we evaluated, including disorders prominent in childhood. This pattern of results indicates that the current classification scheme for major psychiatric disorders does not align well with the underlying genetic basis of these disorders. Currently, only schizophrenia has a sufficient number of genome-wide significant loci to conduct MR analysis, but the bidirectionally significant MR results are consistent with a shared biological basis for major depression and schizophrenia.

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

The dominant psychiatric nosological systems were principally designed for clinical utility and are based on data that emerge during human interactions (i.e., observable signs and reported symptoms) and not objective measurements of pathophysiology. MDD is frequently comorbid with other psychiatric disorders, and the phenotypic comorbidity has an underlying structure that reflects shared origins (as inferred from factor analyses and twin studies)[70–73]. Our genetic results add to this knowledge: major depression is not a discrete entity at any level of analysis. Rather, our data strongly suggest the existence of biological processes common to major depression and schizophrenia (and likely other psychiatric disorders).

Finally, as expected, we found that major depression had modest h^2_{SNP} (8.7%), as it is a complex malady with both genetic and environmental determinants. We found that major depression has a very high genetic correlation with proxy measures that can be briefly assessed. Lifetime major depressive disorder requires a constellation of signs and symptoms whose reliable scoring requires an extended interview with a trained clinician. However, the common variant genetic architecture of lifetime major depression in these seven cohorts (containing many subjects medically treated for MDD) has strong overlap with that of current depressive symptoms in general community samples. Similar relationships of clinically defined ADHD or autism with quantitative genetic variation in the population have been reported[74,75]. The 'disorder-versus-symptom' relationship has been debated extensively[76], but our data indicate that the common variant genetic overlap is very high. This finding has important implications.

One implication is for future genetic studies. In a first phase, it should be possible to elucidate the bulk of the common variant genetic architecture of MDD using a cost-effective shortcut—large studies of genotyped individuals who complete online self-report assessments of lifetime MDD (a sample size approaching 1 million MDD cases may be achievable by 2020). Use of online assessment could allow for recording of a broad range of phenotypes including comorbidities and putative environmental exposures, but with the key feature being large samples with consistently assessed measures. In a second phase, with a relatively complete understanding of the genetic basis of major depression, one could then evaluate smaller samples of carefully phenotyped individuals with MDD to understand the clinical importance of the genetic results. Subsequent empirical studies may show that it is possible to stratify MDD cases at first presentation to identify individuals at high risk for recurrence, poor outcome, poor treatment response, or who might subsequently develop a psychiatric disorder requiring alternative pharmacotherapy (for example, schizophrenia or bipolar disorder). This could form a cornerstone of precision medicine in psychiatry.

In summary, this genome-wide association meta-analysis of 135,438 MDD and major depression cases and 344,901 controls identified 44 loci. An extensive set of companion analyses provide insights into the nature of MDD as well as its neurobiology, therapeutic relevance, and genetic and biological interconnections to other psychiatric disorders. Comprehensive elucidation of these features is the primary goal of our genetic studies of MDD.

URLs. 1000 Genomes Project multi-ancestry imputation panel, https://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html; 23andMe privacy policy, https://www.23andme.com/en-eu/about/privacy; Bedtools, https://bedtools.readthedocs.io; dbGaP, https://www.ncbi.nlm.nih.gov/gap; genotype-based checksums for relatedness determination, http://www.broadinstitute.org/~sripke/share_links/checksums_download; GSMR, http://cnsgenomics.com/software/gsmr/; GTEx, http://www.gtexportal.org/home/datasets; GTMapTool, http://infochim.u-strasbg.fr/mobyle-cgi/portal.py#forms::gtmaptool; LD-Hub, http://ldsc.broadinstitute.org/; PGC, http://www.med.unc.edu/pgc; NIH NeuroBiobank, https://neurobiobank.nih.gov/; PGC ricopili genome-wide association analysis pipeline, https://github.com/Nealelab/ricopili; SMR, http://cnsgenomics.com/software/smr/#Overview; TWAS, http://gusevlab.org/projects/fusion/; UK Biobank, http://www.ukbiobank.ac.uk/.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0090-3.

Received: 28 July 2017; Accepted: 14 February 2018;
Published online: 26 April 2018

References

1. Kessler, R. C. & Bromet, E. J. The epidemiology of depression across cultures. *Annu. Rev. Public Health* **34**, 119–138 (2013).
2. Judd, L. L. The clinical course of unipolar major depressive disorders. *Arch. Gen. Psychiatry* **54**, 989–991 (1997).
3. Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T. & Murray, C. J. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* **367**, 1747–1757 (2006).
4. Wittchen, H. U. et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21**, 655–679 (2011).
5. Ferrari, A. J. et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med.* **10**, e1001547 (2013).
6. Angst, F., Stassen, H. H., Clayton, P. J. & Angst, J. Mortality of patients with mood disorders: follow-up over 34–38 years. *J. Affect. Disord.* **68**, 167–181 (2002).
7. Gustavsson, A. et al. Cost of disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21**, 718–779 (2011).
8. Murray, C. J. et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2197–2223 (2012).
9. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *Am. J. Psychiatry* **157**, 1552–1562 (2000).
10. Rice, F., Harold, G. & Thapar, A. The genetic aetiology of childhood depression: a review. *J. Child Psychol. Psychiatry* **43**, 65–79 (2002).
11. Viktorin, A. et al. Heritability of perinatal depression and genetic overlap with nonperinatal depression. *Am. J. Psychiatry* **73**, 158–165 (2016).
12. Levinson, D. F. et al. Genetic studies of major depressive disorder: why are there no GWAS findings, and what can we do about it. *Biol. Psychiatry* **76**, 510–512 (2014).
13. Major Depressive Disorder Working Group of the PGC. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
14. Hek, K. et al. A genome-wide association study of depressive symptoms. *Biol. Psychiatry* **73**, 667–678 (2013).
15. CONVERGE Consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588–591 (2015).
16. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
17. Sullivan, P. F. et al. Psychiatric genomics. An update and an agenda. *Am. J. Psychiatry* **175**, 15–27 (2018).
18. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
19. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
20. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*. *Nat. Genet.* **43**, 977–983 (2011).
21. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
22. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
23. Wray, N. R. et al. Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol. Psychiatry* **17**, 36–48 (2012).

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

24. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
25. Meier, S. M. et al. High loading of polygenic risk in cases with chronic schizophrenia. *Mol. Psychiatry* **21**, 969–974 (2016).
26. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
27. Wray, N. R. & Maier, R. Genetic basis of complex genetic disease: the contribution of disease heterogeneity to missing heritability. *Curr. Epidemiol. Rep.* **1**, 220–227 (2014).
28. Hyde, C. L. et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).
29. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
30. Berndt, S. I. et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat. Genet.* **45**, 501–512 (2013).
31. Bradfield, J. P. et al. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat. Genet.* **44**, 526–531 (2012).
32. Speliotes, E. K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
33. Willer, C. J. et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).
34. Thorleifsson, G. et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18–24 (2009).
35. Fogel, B. L. et al. RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Hum. Mol. Genet.* **21**, 4171–4186 (2012).
36. Gehman, L. T. et al. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* **43**, 706–711 (2011).
37. Pariante, C. M. & Lightman, S. L. The HPA axis in major depression: classical theories and new developments. *Trends Neurosci.* **31**, 464–468 (2008).
38. Choi, Y. et al. SALM5 trans-synaptically interacts with LAR-RPTPs in a splicing-dependent manner to regulate synapse development. *Sci. Rep.* **6**, 26676 (2016).
39. Mah, W. et al. Selected SALM (synaptic adhesion-like molecule) family proteins regulate synapse formation. *J. Neurosci.* **30**, 5559–5568 (2010).
40. Zhu, Y. et al. Neuron-specific SALM5 limits inflammation in the CNS via its interaction with HVEM. *Sci. Adv.* **2**, e1500637 (2016).
41. Amiel, J. et al. Mutations in *TCF4*, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt–Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am. J. Hum. Genet.* **80**, 988–993 (2007).
42. Akbarian, S. et al. The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).
43. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
44. Schmaal, L. et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol. Psychiatry* **22**, 900–909 (2017).
45. Cahoy, J. D. et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
46. Finucane, H. K. et al. Partitioning heritability by functional category using GWAS summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
47. Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
48. Simonti, C. N. et al. The phenotypic legacy of admixture between modern humans and Neandertals. *Science* **351**, 737–741 (2016).
49. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
50. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
51. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
52. Smemo, S. et al. Obesity-associated variants within *FTO* form long-range functional connections with *IRX3*. *Nature* **507**, 371–375 (2014).
53. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
54. Martin, J. S. et al. HUGIn. Hi-C Unifying Genomic Interrogator. *Bioinformatics* **33**, 3793–3795 (2017).
55. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18**, 199–209 (2015).
56. De Rubeis, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
57. Genovese, G. et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).
58. Gaspar, H. A. & Breen, G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. *Sci. Rep.* **7**, 12460 (2017).
59. Breen, G. et al. Translating genome-wide association findings into new therapeutics for psychiatry. *Nat. Neurosci.* **19**, 1392–1396 (2016).
60. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
61. Sniekers, S. et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* **49**, 1107–1112 (2017).
62. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
63. Nikpay, M. et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
64. Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).
65. Wray, N. R., Lee, S. H. & Kendler, K. S. Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. *Eur. J. Hum. Genet.* **20**, 668–674 (2012).
66. Hippocrates. *The Aphorisms of Hippocrates* (Collins & Co., New York, 1817).
67. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* (in the press).
68. Yang, X. et al. Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* **164**, 805–817 (2016).
69. Zhang, X. et al. Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. *Cell* **166**, 1147–1162 (2016).
70. Kessler, R. C. et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *J. Am. Med. Assoc.* **289**, 3095–3105 (2003).
71. Hasin, D. S., Goodwin, R. D., Stinson, F. S. & Grant, B. F. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Arch. Gen. Psychiatry* **62**, 1097–1106 (2005).
72. Kendler, K. S. et al. The structure of genetic and environmental risk factors for syndromal and subsyndromal common DSM-IV axis I and all axis II disorders. *Am. J. Psychiatry* **168**, 29–39 (2011).
73. Kendler, K. S., Prescott, C. A., Myers, J. & Neale, M. C. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch. Gen. Psychiatry* **60**, 929–937 (2003).
74. Robinson, E. B. et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555 (2016).
75. Middeldorp, C. M. et al. A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts. *J. Am. Acad. Child. Adolesc. Psychiatry* **55**, 896–905 (2016).
76. Kendell, R. E. The classification of depressions: a review of contemporary confusion. *Br. J. Psychiatry* **129**, 15–28 (1976).
77. Verduijn, J. et al. Using clinical characteristics to identify which patients with major depressive disorder have a higher genetic load for three psychiatric disorders. *Biol. Psychiatry* **81**, 316–324 (2017).
78. Smith, B. H. et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* **42**, 689–700 (2013).
79. Fernandez-Pujals, A. M. et al. Epidemiology and heritability of major depressive disorder, stratified by age of onset, sex, and illness course in Generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* **10**, e0142197 (2015).
80. Banda, Y. et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Genetics* **200**, 1285–1295 (2015).
81. Pedersen, C. B. et al. The iPSYCH2012 case–cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2018).
82. Allen, N. E., Sudlow, C., Peakman, T., Collins, R. & UK Biobank. UK Biobank data: come and get it. *Sci. Transl. Med.* **6**, 224ed4 (2014).
83. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for ADHD. Preprint at *bioRxiv* https://www.biorxiv.org/content/early/2017/06/03/145581 (2017).

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

84. Boraska, V. et al. A genome-wide association study of anorexia nervosa. *Mol. Psychiatry* **19**, 1085–1094 (2014).
85. Otowa, T. et al. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol. Psychiatry* **21**, 1391–1399 (2016).
86. Grove, J. et al. Common risk variants identified in autism spectrum disorder. Preprint at *bioRxiv* https://www.biorxiv.org/content/early/2017/11/27/224774 (2017).
87. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
88. Deary, V. et al. Genetic contributions to self-reported tiredness. *Mol. Psychiatry* **23**, 609–620 (2018).
89. Rietveld, C. A. et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).
90. Lu, Y. et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat. Commun.* **7**, 10495 (2016).
91. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
92. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
93. Perry, J. R. et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
94. Barban, N. et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016).
95. Pilling, L. C. et al. Human longevity is influenced by many genetic variants: evidence from 75,000 UK Biobank participants. *Aging* **8**, 547–560 (2016).
96. Patel, Y. M. et al. Novel association of genetic markers affecting CYP2A6 activity and lung cancer risk. *Cancer Res.* **76**, 5768–5776 (2016).

Acknowledgements

Full acknowledgments are in the Supplementary Note. We are deeply indebted to the investigators who comprise the PGC, and to the hundreds of thousands of subjects who have shared their life experiences with PGC investigators. A full list of funding is in the Supplementary Note. Major funding for the PGC is from the US National Institutes of Health (U01 MH109528 and U01 MH109532). Statistical analyses were carried out on the NL Genetic Cluster Computer (http://www.geneticcluster.org/) hosted by SURFsara. The iPSYCH team acknowledges funding from the Lundbeck Foundation (grants R102-A9118 and R155-2014-1724), the Stanley Medical Research Institute, the European Research Council (project 294838), the Novo Nordisk Foundation for supporting the Danish National Biobank resource, and Aarhus and Copenhagen Universities and University Hospitals, including support to the iSEQ Center, the GenomeDK HPC facility, and the CIRRAU Center. This research has been conducted using the UK Biobank Resource (see URLs), including applications 4844 and 6818. Finally, we thank the members of the eQTLGen Consortium for allowing us to use their very large eQTL database ahead of publication. Its members are listed in Supplementary Table 14.

Some data used in this study were obtained from dbGaP (see URLs). dbGaP accession phs000021: funding support for the Genome-Wide Association of Schizophrenia Study was provided by the National Institute of Mental Health (R01 MH67257, R01 MH59588, R01 MH59571, R01 MH59565, R01 MH59587, R01 MH60870, R01 MH59566, R01 MH59586, R01 MH61675, R01 MH60879, R01 MH81800, U01 MH46276, U01 MH46289, U01 MH46318, U01 MH79469, and U01 MH79470), and the genotyping of samples was provided through the Genetic Association Information Network (GAIN). Samples and associated phenotype data for the Genome-Wide Association of Schizophrenia Study were provided by the Molecular Genetics of Schizophrenia Collaboration (principal investigator P. V. Gejman, Evanston Northwestern Healthcare (ENH) and Northwestern University, Evanston, IL, USA). dbGaP accession phs000196: this work used in part data from the NINDS dbGaP database from the CIDR:NGRC PARKINSON'S DISEASE STUDY. dbGaP accession phs000187: High-Density SNP Association Analysis of Melanoma: Case–Control and Outcomes Investigation. Research support to collect data and develop an application to support this project was provided by P50 CA093459, P50 CA097007, R01 ES011740, and R01 CA133996 from the NIH.

Author contributions

Writing group: G.B., A.D.B., D.F.L., C.M.L., S.R., P.F.S., N.R.W. **PGC MDD PI group:** V.A., B.T.B., K.B., D.I.B., G.B., A.D.B., S.C., U.D., J.R.D., E.D., K.D., T.E., E.J.C.d.G., H.J.G., S.P.H., C. Hayward, A.C.H., D.M.H., K.S.K., S.K., D.F.L., C.M.L., G.L., Q.S.L., S.L., P.A.F.M., P.K.M., N.G.M., A.M.M., A.M., O.M., P.B.M., B.M.-M., M. Nordentoft, M.M.N., M.C.O'D., S.A.P., N.L.P., B.W.P., R.H.P., D.J.P., J.B.P., M.P., M. Rietschel, C.S., T.G.S., J.W.S., K.S., P.F.S., H. Tiemeier, R.U., H.V., M.M.W., T.W., A.R.W., N.R.W. **Bioinformatics:** 23andMe Research Team, M.J.A., S.V.d.A., G.B., J.B., A.D.B., E.C., J.H.C., T.-K.C., J.R.I.C., L.C.-C., eQTLGen Consortium, G.E.C., C.A.C., G.D., E.M.D., T.E., A.J.F., H.A.G., P.G.-R., J.G., L.S.H., E.H., T.F.H., C. Hayward, M.H., R.J., F.J., Z.K., Q.S.L., Yihan Li, P.A.L., X.L., L.L., D.J.M., S.E.M., E.M., Y.M., J. Mill, J.N.P., B.W.P., W.J.P., G.P., P.Q., L.S., S.I.S., C.A.S., P.F.S., K.E.T., A.T., P.A.T., A.G.U., Y. Wang, S.M.W., N.R.W., H.S.X. **Clinical:** E.A., T.M.A., V.A., B.T.B., A.T.F.B., K.B., E.B.B., D.H.R.B., H.N.B., A.D.B., N. Craddock, U.D., J.R.D., N.D., K.D., M.G., F.S.G., H.J.G., A.C.H., A.M.v.H., I.B.H., M.I., S.K., J. Krogh, D.F.L., S.L., D.J.M., D.F.M., P.A.F.M., W.M., N.G.M., P. McGrath, P. McGuffin, A.M.M., A.M., C.M.M., S.S.M., F.M.M., O.M., P.B.M., D.R.N., H.O., M.J.O., C.B.P., M.G.P., J.B.P., J.A.Q., J.P.R., M. Rietschel, C.S., R. Schoevers, E.S., G.C.B.S., D.J.S., F.S., J. Strohmaier, D.U., M.M.W., J.W., T.W., G.W. **Genomic assays:** G.B., H.N.B., J.B.-G., M.B.-H., A.D.B., S.C., T.-K.C., F.D., A.J.F., S.P.H., C.S.H., A.C.H., P.H., G.H., C. Horn, J.A.K., P.A.F.M., L.M., G.W.M., M. Nauck, M.M.N., M. Rietschel, M. Rivera, E.C.S., T.G.S., S.I.S., H.S., F.S., T.E.T., J.T., A.G.U., S.H.W. **Obtained funding for primary MDD samples:** B.T.B., K.B., D.H.R.B., D.I.B., G.B., H.N.B., A.D.B., S.C., J.R.D., I.J.D., E.D., T.C.E., T.E., H.J.G., S.P.H., A.C.H., D.M.H., I.S.K., D.F.L., C.M.L., G.L., Q.S.L., S.L., P.A.F.M., W.M., N.G.M., P. McGuffin, A.M.M., A.M., G.W.M., O.M., P.B.M., M. Nordentoft, D.R.N., M.M.N., P.F.O'R., B.W.P., D.J.P., J.B.P., M.P., M. Rietschel, C.S., T.G.S., G.C.B.S., J.H.S., D.J.S., H.S., K.S., P.F.S., T.E.T., H. Tiemeier, A.G.U., H.V., M.M.W., T.W., N.R.W. **Statistical analysis:** 23andMe Research Team, A.A., M.J.A., T.F.M.A., S.V.d.A., S.-A.B., K.B., T.B.B., G.B., E.M.B., A.D.B., N. Cai, T.-K.C., J.R.I.C., B.C.-D., H.S.D., G.D., N.D., C.V.D., E.C.D., N.E., V.E.-P., T.E., H.K.F., J.F., H.A.G., S.D.G., J.G., L.S.H., C. Hayward, A.C.H., S.H., D.A.H., J.-J.H., C.L.H., M.I., E.J., F.F.H.K., J. Kraft, W.W.K., Z.K., J.M.L., C.M.L., Q.S.L., Yun Li, D.J.M., P.A.F.M., R.M.M., J. Marchini, M.M., H.M., A.M.M., S.E.M., D.M., E.M., Y.M., S.S.M., S.M., N.M., B.M.-M., B.N., M.G.N., D.R.N., P.F.O'R., R.E.P., E.P., W.J.P., G.P., D.P., S.M.P., B.P.R., S.R., M. Rivera, R. Saxena, C.S., L.S., J. Shi, S.I.S., H.S., S.S., P.F.S., K.E.T., H. Teismann, A.T., W.T., P.A.T., T.E.T., C.T., M. Traylor, V.T., M. Trzaskowski, A.V., P.M.V., Y. Wang, B.T.W., J.W., T.W., N.R.W., Y. Wu, J.Y., F.Z.

Competing interests

A.T.F.B. is on speaker's bureaus for Lundbeck and GlaxoSmithKline. G.C. is a cofounder of Element Genomics. E.D. was an employee of Hoffmann–La Roche at the time this study was conducted and a consultant to Roche and Pierre-Fabre. N.E. is employed by 23andMe, Inc., and owns stock in 23andMe, Inc. D.H. is an employee of and owns stock options in 23andMe, Inc. S.P. is an employee of Pfizer, Inc. C.L.H. is an employee of Pfizer, Inc. A.R.W. was a former employee and stockholder of Pfizer, Inc. J.A.Q. was an employee of Hoffmann–La Roche at the time this study was conducted. H.S. is an employee of deCODE Genetics/Amgen. K.S. is an employee of deCODE Genetics/Amgen. S.S. is an employee of deCODE Genetics/Amgen. P.F.S. is on the scientific advisory board for Pfizer, Inc., and the advisory committee for Lundbeck. T.E.T. is an employee of deCODE Genetics/Amgen. C.T. is an employee of and owns stock options in 23andMe, Inc.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0090-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to N.R.W. or P.F.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

Naomi R. Wray[1,2,160*], Stephan Ripke[3,4,5,160], Manuel Mattheisen[6,7,8,9,160], Maciej Trzaskowski[1,160], Enda M. Byrne[1], Abdel Abdellaoui[10], Mark J. Adams[11], Esben Agerbo[8,12,13], Tracy M. Air[14], Till M. F. Andlauer[15,16], Silviu-Alin Bacanu[17], Marie Bækvad-Hansen[8,18], Aartjan F. T. Beekman[19], Tim B. Bigdeli[17,20], Elisabeth B. Binder[15,21], Douglas R. H. Blackwood[11], Julien Bryois[22], Henriette N. Buttenschøn[7,8,23], Jonas Bybjerg-Grauholm[8,18], Na Cai[24,25], Enrique Castelao[26], Jane Hvarregaard Christensen[6,7,8], Toni-Kim Clarke[11], Jonathan I. R. Coleman[27], Lucía Colodro-Conde[28], Baptiste Couvy-Duchesne[2,29], Nick Craddock[30], Gregory E. Crawford[31,32], Cheynna A. Crowley[33], Hassan S. Dashti[3,34], Gail Davies[35], Ian J. Deary[35], Franziska Degenhardt[36,37], Eske M. Derks[28], Nese Direk[38,39], Conor V. Dolan[10], Erin C. Dunn[40,41,42], Thalia C. Eley[27], Nicholas Eriksson[43], Valentina Escott-Price[44], Farnush Hassan Farhadi Kiadeh[45], Hilary K. Finucane[46,47], Andreas J. Forstner[36,37,48,49], Josef Frank[50], Héléna A. Gaspar[27], Michael Gill[51], Paola Giusti-Rodríguez[52], Fernando S. Goes[53], Scott D. Gordon[54], Jakob Grove[6,7,8,55], Lynsey S. Hall[11,56], Eilis Hannon[57], Christine Søholm Hansen[8,18], Thomas F. Hansen[58,59,60], Stefan Herms[36,37,49], Ian B. Hickie[61], Per Hoffmann[36,37,49], Georg Homuth[62], Carsten Horn[63], Jouke-Jan Hottenga[10], David M. Hougaard[8,18], Ming Hu[64], Craig L. Hyde[65], Marcus Ising[66], Rick Jansen[19], Fulai Jin[67,68], Eric Jorgenson[69], James A. Knowles[70], Isaac S. Kohane[71,72,73], Julia Kraft[5], Warren W. Kretzschmar[74], Jesper Krogh[75], Zoltán Kutalik[76,77], Jacqueline M. Lane[3,34,78], Yihan Li[74], Yun Li[33,52], Penelope A. Lind[28], Xiaoxiao Liu[68], Leina Lu[68], Donald J. MacIntyre[79,80], Dean F. MacKinnon[53], Robert M. Maier[2], Wolfgang Maier[81], Jonathan Marchini[82], Hamdi Mbarek[10], Patrick McGrath[83], Peter McGuffin[27], Sarah E. Medland[28], Divya Mehta[2,84], Christel M. Middeldorp[10,85,86], Evelin Mihailov[87], Yuri Milaneschi[19], Lili Milani[87], Jonathan Mill[57], Francis M. Mondimore[53], Grant W. Montgomery[1], Sara Mostafavi[88,89], Niamh Mullins[27], Matthias Nauck[90,91], Bernard Ng[89], Michel G. Nivard[10], Dale R. Nyholt[92], Paul F. O'Reilly[27], Hogni Oskarsson[93], Michael J. Owen[94], Jodie N. Painter[28], Carsten Bøcker Pedersen[8,12,13], Marianne Giørtz Pedersen[8,12,13], Roseann E. Peterson[17,95], Erik Pettersson[22], Wouter J. Peyrot[19], Giorgio Pistis[26], Danielle Posthuma[96,97], Shaun M. Purcell[98], Jorge A. Quiroz[99], Per Qvist[6,7,8], John P. Rice[100], Brien P. Riley[17], Margarita Rivera[27,101], Saira Saeed Mirza[39], Richa Saxena[3,34,78], Robert Schoevers[102], Eva C. Schulte[103,104], Ling Shen[69], Jianxin Shi[105], Stanley I. Shyn[106], Engilbert Sigurdsson[107], Grant B. C. Sinnamon[108], Johannes H. Smit[19], Daniel J. Smith[109], Hreinn Stefansson[110], Stacy Steinberg[110], Craig A. Stockmeier[111], Fabian Streit[50], Jana Strohmaier[50], Katherine E. Tansey[112], Henning Teismann[113], Alexander Teumer[114], Wesley Thompson[8,59,115,116], Pippa A. Thomson[117], Thorgeir E. Thorgeirsson[110], Chao Tian[43], Matthew Traylor[118], Jens Treutlein[50], Vassily Trubetskoy[5], André G. Uitterlinden[119], Daniel Umbricht[120], Sandra Van der Auwera[121], Albert M. van Hemert[122], Alexander Viktorin[22], Peter M. Visscher[1,2], Yunpeng Wang[8,59,115], Bradley T. Webb[123], Shantel Marie Weinsheimer[8,59], Jürgen Wellmann[113], Gonneke Willemsen[10], Stephanie H. Witt[50], Yang Wu[1], Hualin S. Xi[124], Jian Yang[1,2], Futao Zhang[1], eQTLGen[125], 23andMe, Volker Arolt[126], Bernhard T. Baune[14], Klaus Berger[113], Dorret I. Boomsma[10], Sven Cichon[36,49,127,128], Udo Dannlowski[126], E. C. J. de Geus[10,129], J. Raymond DePaulo[53], Enrico Domenici[130], Katharina Domschke[131], Tõnu Esko[3,87], Hans J. Grabe[121], Steven P. Hamilton[132], Caroline Hayward[133], Andrew C. Heath[100], David A. Hinds[43], Kenneth S. Kendler[17], Stefan Kloiber[66,134,135], Glyn Lewis[136], Qingqin S. Li[137], Susanne Lucae[66], Pamela F. A. Madden[100], Patrik K. Magnusson[22], Nicholas G. Martin[54], Andrew M. McIntosh[11,35], Andres Metspalu[87,138], Ole Mors[8,139], Preben Bo Mortensen[7,8,12,13], Bertram Müller-Myhsok[15,16,140], Merete Nordentoft[8,141], Markus M. Nöthen[36,37], Michael C. O'Donovan[94], Sara A. Paciga[142], Nancy L. Pedersen[22], Brenda W. J. H. Penninx[19], Roy H. Perlis[41,143], David J. Porteous[117],

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

James B. Potash[144], Martin Preisig[26], Marcella Rietschel[50], Catherine Schaefer[69], Thomas G. Schulze[50,53,104,145,146], Jordan W. Smoller[40,41,42], Kari Stefansson[110,147], Henning Tiemeier[39,148,149], Rudolf Uher[150], Henry Völzke[114], Myrna M. Weissman[83,151], Thomas Werge[8,59,152], Ashley R. Winslow[153,154], Cathryn M. Lewis[27,155,160], Douglas F. Levinson[156,160], Gerome Breen[27,157,160], Anders D. Børglum[6,7,8,160], Patrick F. Sullivan[22,52,158,160]* and the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium

[1]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [2]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [3]Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [4]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [5]Department of Psychiatry and Psychotherapy, Universitätsmedizin Berlin Campus Charité Mitte, Berlin, Germany. [6]Department of Biomedicine, Aarhus University, Aarhus, Denmark. [7]iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark. [8]iPSYCH, Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark. [9]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [10]Department of Biological Psychology and EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [11]Division of Psychiatry, University of Edinburgh, Edinburgh, UK. [12]Centre for Integrated Register-Based Research, Aarhus University, Aarhus, Denmark. [13]National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [14]Discipline of Psychiatry, University of Adelaide, Adelaide, South Australia, Australia. [15]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. [16]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [17]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA. [18]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [19]Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, The Netherlands. [20]Virginia Institute for Psychiatric and Behavior Genetics, Richmond, VA, USA. [21]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA. [22]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [23]Department of Clinical Medicine, Translational Neuropsychiatry Unit, Aarhus University, Aarhus, Denmark. [24]Statistical Genomics and Systems Genetics, European Bioinformatics Institute (EMBL-EBI), Cambridge, UK. [25]Human Genetics, Wellcome Trust Sanger Institute, Cambridge, UK. [26]Department of Psychiatry, University Hospital of Lausanne, Prilly, Switzerland. [27]MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, UK. [28]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia. [29]Centre for Advanced Imaging, University of Queensland, Brisbane, Queensland, Australia. [30]Psychological Medicine, Cardiff University, Cardiff, UK. [31]Center for Genomic and Computational Biology, Duke University, Durham, NC, USA. [32]Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC, USA. [33]Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [34]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [35]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [36]Institute of Human Genetics, University of Bonn, Bonn, Germany. [37]Life & Brain Center, Department of Genomics, University of Bonn, Bonn, Germany. [38]Psychiatry, Dokuz Eylul University School of Medicine, Izmir, Turkey. [39]Epidemiology, Erasmus MC, Rotterdam, The Netherlands. [40]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA. [41]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [42]Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, USA. [43]Research, 23andMe, Inc., Mountain View, CA, USA. [44]Neuroscience and Mental Health, Cardiff University, Cardiff, UK. [45]Bioinformatics, University of British Columbia, Vancouver, British Columbia, Canada. [46]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [47]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA. [48]Department of Psychiatry (UPK), University of Basel, Basel, Switzerland. [49]Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland. [50]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [51]Department of Psychiatry, Trinity College Dublin, Dublin, Ireland. [52]Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [53]Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD, USA. [54]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [55]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [56]Institute of Genetic Medicine, Newcastle University, Newcastle-upon-Tyne, UK. [57]University of Exeter Medical School, Exeter, UK. [58]Danish Headache Centre, Department of Neurology, Rigshospitalet, Glostrup, Denmark. [59]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region of Denmark, Copenhagen, Denmark. [60]iPSYCH, Lundbeck Foundation Initiative for Psychiatric Research, Copenhagen, Denmark. [61]Brain and Mind Centre, University of Sydney, Sydney, New South Wales, Australia. [62]Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University Medicine and Ernst Moritz Arndt University Greifswald, Greifswald, Germany. [63]Roche Pharmaceutical Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, F. Hoffmann–La Roche, Ltd, Basel, Switzerland. [64]Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA. [65]Statistics, Pfizer Global Research and Development, Groton, CT, USA. [66]Max Planck Institute of Psychiatry, Munich, Germany. [67]Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, USA. [68]Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, USA. [69]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. [70]Psychiatry and Behavioral Sciences, University of Southern California, Los Angeles, CA, USA. [71]Informatics Program, Boston Children's Hospital, Boston, MA, USA. [72]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [73]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [74]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [75]Department of Endocrinology at Herlev University Hospital, University of Copenhagen, Copenhagen, Denmark. [76]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [77]Institute of Social and Preventive Medicine (IUMSP), University Hospital of Lausanne, Lausanne, Switzerland. [78]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA, USA. [79]Mental Health, NHS 24 Glasgow, UK. [80]Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. [81]Department of Psychiatry and Psychotherapy, University of Bonn, Bonn, Germany. [82]Statistics, University of Oxford, Oxford, UK. [83]Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, USA. [84]School of Psychology and Counseling, Queensland University of Technology, Brisbane, Queensland, Australia. [85]Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, South Brisbane, Queensland, Australia. [86]Child Health Research Centre, University of Queensland, Brisbane, Queensland, Australia. [87]Estonian Genome Center, University of Tartu, Tartu, Estonia. [88]Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada. [89]Statistics, University of British Columbia, Vancouver, British Columbia, Canada. [90]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, University Medicine, University Medicine Greifswald, Greifswald, Germany. [91]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [92]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia.

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

[93]Humus, Reykjavik, Iceland. [94]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK. [95]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [96]Complex Trait Genetics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [97]Clinical Genetics, Vrije Universiteit Medical Center, Amsterdam, The Netherlands. [98]Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, USA. [99]Solid Biosciences, Boston, MA, USA. [100]Department of Psychiatry, Washington University in St. Louis School of Medicine, St. Louis, MO, USA. [101]Department of Biochemistry and Molecular Biology II, Institute of Neurosciences, Center for Biomedical Research, University of Granada, Granada, Spain. [102]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. [103]Department of Psychiatry and Psychotherapy, Medical Center of the University of Munich, Campus Innenstadt, Munich, Germany. [104]Institute of Psychiatric Phenomics and Genomics (IPPG), Medical Center of the University of Munich, Campus Innenstadt, Munich, Germany. [105]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. [106]Behavioral Health Services, Kaiser Permanente Washington, Seattle, WA, USA. [107]Faculty of Medicine, Department of Psychiatry, University of Iceland, Reykjavik, Iceland. [108]School of Medicine and Dentistry, James Cook University, Townsville, Queensland, Australia. [109]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK. [110]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [111]Psychiatry and Human Behavior, University of Mississippi Medical Center, Jackson, MS, USA. [112]College of Biomedical and Life Sciences, Cardiff University, Cardiff, UK. [113]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [114]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [115]KG Jebsen Centre for Psychosis Research, Norway Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [116]Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA. [117]Medical Genetics Section, CGEM, IGMM, University of Edinburgh, Edinburgh, UK. [118]Clinical Neurosciences, University of Cambridge, Cambridge, UK. [119]Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. [120]Roche Pharmaceutical Research and Early Development, Neuroscience, Ophthalmology and Rare Diseases Discovery and Translational Medicine Area, Roche Innovation Center Basel, F. Hoffmann–La Roche, Ltd, Basel, Switzerland. [121]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [122]Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands. [123]Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [124]Computational Sciences Center of Emphasis, Pfizer Global Research and Development, Cambridge, MA, USA. [125]A full list of members and affiliations appears in Supplementary Table 14. [126]Department of Psychiatry, University of Münster, Munster, Germany. [127]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany. [128]Institute of Medical Genetics and Pathology, University Hospital Basel, University of Basel, Basel, Switzerland. [129]Amsterdam Public Health Institute, Vrije Universiteit Medical Center, Amsterdam, The Netherlands. [130]Centre for Integrative Biology, Università degli Studi di Trento, Trento, Italy. [131]Department of Psychiatry and Psychotherapy, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany. [132]Psychiatry, Kaiser Permanente Northern California, San Francisco, CA, USA. [133]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [134]Centre for Addiction and Mental Health, Toronto, Ontario, Canada. [135]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada. [136]Division of Psychiatry, University College London, London, UK. [137]Neuroscience Therapeutic Area, Janssen Research and Development, LLC, Titusville, NJ, USA. [138]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [139]Psychosis Research Unit, Aarhus University Hospital, Risskov, Aarhus, Denmark. [140]Institute of Translational Medicine, University of Liverpool, Liverpool, UK. [141]Mental Health Center Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark. [142]Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Groton, CT, USA. [143]Psychiatry, Harvard Medical School, Boston, MA, USA. [144]Psychiatry, University of Iowa, Iowa City, IA, USA. [145]Human Genetics Branch, NIMH Division of Intramural Research Programs, Bethesda, MD, USA. [146]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany. [147]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [148]Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, The Netherlands. [149]Psychiatry, Erasmus MC, Rotterdam, The Netherlands. [150]Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada. [151]Division of Epidemiology, New York State Psychiatric Institute, New York, NY, USA. [152]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [153]Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Cambridge, MA, USA. [154]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [155]Department of Medical and Molecular Genetics, King's College London, London, UK. [156]Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [157]NIHR BRC for Mental Health, King's College London, London, UK. [158]Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [159]These authors contributed equally: Naomi R. Wray, Stephan Ripke, Manuel Mattheisen, Maciej Trzaskowski. [160]These authors jointly directed this work: Cathryn M. Lewis, Douglas F. Levinson, Gerome Breen, Anders D. Børglum, Patrick F. Sullivan. *e-mail: naomi.wray@uq.edu.au; pfsulliv@med.unc.edu

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

Methods

PGC29 cohort. Our analysis was anchored in a genome-wide association mega-analysis of 29 samples of European ancestry (16,823 MDD cases and 25,632 controls). Supplementary Table 1 summarizes the source and inclusion/exclusion criteria for cases and controls for each sample. All PGC29 samples passed a structured methodological review by MDD assessment experts (D.F.L. and K.S.K.). Cases were required to meet international consensus criteria (DSM-IV, ICD-9, or ICD-10)[97–99] for a lifetime diagnosis of MDD established using structured diagnostic instruments from assessments by trained interviewers, clinician-administered checklists, or medical record review. All cases met standard criteria for MDD, were directly interviewed (28/29 samples), or had medical record review by an expert diagnostician (1/29 samples), and most were ascertained from clinical sources (19/29 samples). Controls in most samples were screened for the absence of lifetime MDD (22/29 samples) and randomly selected from the population.

Additional cohorts. We critically evaluated six independent European-ancestry cohorts (118,635 cases and 319,269 controls). Supplementary Table 2 summarizes the source and inclusion/exclusion criteria for cases and controls for each cohort. These cohorts used a range of methods for assessing MDD or major depression. Most studies included here applied otherwise typical inclusion and exclusion criteria for both cases and controls (for example, excluding cases with lifetime bipolar disorder or schizophrenia and excluding controls with major depression).

Cohort comparability. Supplementary Table 3 summarizes the numbers of cases and controls in PGC29 and the six additional cohorts. The most direct and important way to evaluate the comparability of these cohorts for a genome-wide association meta-analysis is using SNP genotype data[22,24]. We used LD score (LDSC) regression (described below) to estimate h^2_{SNP} for each cohort (Supplementary Fig. 1 and Supplementary Table 3) and r_g for all pairwise combinations of the cohorts (Supplementary Table 3b) and to demonstrate no evidence of sample overlap. We used leave-one-sample-out GRSs, finding significant differences in case–control GRS distributions of the left-out sample for all but one PGC29 sample (Supplementary Table 4). For full details of the cohort comparability analyses including GRS analyses, see the Supplementary Note. In GRS analyses, the discovery sample was the genome-wide association sample that provided the allelic weightings for each SNP, used to generate a sum score for each individual in the independent target sample.

Genotyping and quality control. Genotyping procedures can be found in the primary reports for each cohort (summarized in Supplementary Table 3). Individual genotype data for all PGC29 samples, GERA, and iPSYCH were processed using the PGC ricopili pipeline (see URLs) for standardized quality control, imputation, and analysis[19]. The cohorts from deCODE, Generation Scotland, UK Biobank, and 23andMe were processed by the collaborating research teams using comparable procedures. SNPs and insertion–deletion polymorphisms were imputed using the 1000 Genomes Project multi-ancestry reference panel (see URLs)[100]. More detailed information on sample quality control is provided in the Supplementary Note.

LD score regression. LDSC was used to estimate h^2_{SNP} from genome-wide association summary statistics. Estimates of h^2_{SNP} on the liability scale depend on the assumed lifetime prevalence of MDD in the population (K), and we assumed $K = 0.15$ but also evaluated a range of estimates of K to explore sensitivity, including 95% confidence intervals (Supplementary Fig. 1). LDSC bivariate genetic correlations attributable to genome-wide SNPs (r_g) were estimated across all MDD and major depression cohorts and between the full cohort subjected to meta-analysis and other traits and disorders.

LDSC was also used to partition h^2_{SNP} by genomic features[24,46]. We tested for enrichment of h^2_{SNP} based on genomic annotations, partitioning h^2_{SNP} proportional to the base-pair length represented by each annotation. We used the 'baseline model', which consists of 53 functional categories. The categories are fully described elsewhere[46] and included conserved regions[47], USCC gene models (exons, introns, promoters, UTRs), and functional genomic annotations constructed using data from ENCODE[101] and the Roadmap Epigenomics Consortium[102]. We complemented these annotations by adding introgressed regions from the Neanderthal genome in European populations[103] and open chromatin regions from the brain dorsolateral prefrontal cortex. The open chromatin regions were obtained from an ATAC–seq experiment performed in 288 samples ($n = 135$ controls, 137 schizophrenia cases, 10 bipolar cases, and 6 affective disorder cases)[104]. Peaks called with MACS[105] (1% FDR) were retained if their coordinates overlapped in at least two samples. The peaks were recentered and set to a fixed width of 300 bp using the diffbind R package[106]. To prevent upward bias in heritability enrichment estimation, we added two categories created by expanding both the Neanderthal introgressed regions and open chromatin regions by 250 bp on each side.

We used LDSC to estimate r_g between major depression and a range of other disorders, diseases, and human traits[22]. The intent of these comparisons was to evaluate the extent of shared common variant genetic architectures to suggest hypotheses about the fundamental genetic basis of major depression (given its extensive comorbidity with psychiatric and medical conditions and its association with anthropometric and other risk factors). Subject overlap of itself does not bias r_g. These r_g values were mostly based on studies of independent subjects, and the estimates should be unbiased by confounding of genetic and non-genetic effects (except if there is genotype by environment correlation). When GWAS include overlapping samples, r_g remains unbiased but the intercept of the LDSC regression is an estimate of the correlation between the association statistics attributable to sample overlap. These calculations were done using the internal PGC genome-wide association library and with LD-Hub (see URLs)[60].

Integration of genome-wide association findings to tissue and cellular gene expression. We used partitioned LDSC to evaluate which somatic tissues were enriched for major depression heritability[107]. Gene expression data generated using mRNA-seq from multiple human tissues were obtained from GTEx v6p (see URLs). Genes for which <4 samples had at least one read count per million were discarded, and samples with <100 genes with at least one read count per million were excluded. The data were normalized, and a t statistic was obtained for each tissue by comparing the expression in each tissue with the expression of all other tissues with the exception of tissues related to the tissue of interest (for example, brain cortex versus all other tissues excluding other brain samples), using sex and age as covariates. A t statistic was also obtained for each tissue among its related tissues (for example, cortex versus all other brain tissues) to test which brain region was the most associated with major depression, also using sex and age as covariates. The top 10% of the genes with the most extreme t statistic were defined as tissue specific. The coordinates for these genes were extended by a 100-kb window and tested using LD score regression. Significance was obtained from the coefficient z score, which corrects for all other categories in the baseline model.

Lists of genes specifically expressed in neurons, astrocytes, and oligodendrocytes were obtained from Cahoy et al.[45]. As these experiments were done in mice, genes were mapped to human orthologous genes using Ensembl. The coordinates for these genes were extended by a 100-kb window and tested using LD score regression as for the GTEx tissue-specific genes.

We conducted eQTL lookups of the most associated SNPs in each region and report genome-wide association SNPs in LD ($r^2 > 0.8$) with the top eQTLs in the following datasets: eQTLGen Consortium (Illumina arrays in whole blood, $n = 14{,}115$), BIOS (RNA-seq in whole blood, $n = 2{,}116$)[108], NESDA/NTR (Affymetrix arrays in whole blood, $n = 4{,}896$)[109], GEUVADIS (RNA-seq in LCLs, $n = 465$)[110], Rosmap (RNA-seq in cortex, $n = 494$)[111], GTEx (RNA-seq in 44 tissues, $n > 70$)[43], and Common Mind Consortium (CMC; prefrontal cortex, Sage Synapse accession syn5650509, $n = 467$)[51].

We used SMR[49] to identify loci with strong evidence of causality via gene expression and DNA methylation (eQTLs and meQTLs). SMR analysis is limited to significant cis-SNP expression (FDR < 0.05) and SNPs with MAF >0.01 at a Bonferroni-corrected pSMR. Owing to LD, multiple SNPs may be associated with the expression of a gene, and some SNPs are associated with the expression of more than one gene. The aim of SMR is to prioritize variants and genes for subsequent studies, and a test for heterogeneity excludes regions that may harbor multiple causal loci (pHET < 0.05; a very conservative threshold). SMR analyses were conducted using eQTLs from the eQTLGen Consortium (whole blood), GTEx (11 brain tissues), and the CMC[43,51] as well as meQTLs from whole blood[112].

We conducted a transcriptome-wide association study[50] using precomputed expression reference weights for CMC data (5,420 genes with significant cis-SNP heritability) provided with the TWAS/FUSION software. The significance threshold was 0.05/5,420.

DNA looping using Hi-C. Dorsolateral prefrontal cortex (Brodmann area 9) was dissected from post-mortem samples from three adults of European ancestry (C.S.). Cerebra from three fetal brains were obtained from the NIH NeuroBiobank (see URLs; gestation age 17–19 weeks, African ancestry). We used 'easy Hi-C' to assess DNA chromatin (looping) interactions (Supplementary Note).

Gene-wise and pathway analyses. Our approach was guided by rigorous method comparisons conducted by PGC members[55,113]. P values quantifying the degree of association of genes and gene sets with MDD were generated using MAGMA (v1.06)[114]. MAGMA uses Brown's method to combine SNP P values and account for LD. We used Ensembl gene models for 19,079 genes, giving a Bonferroni-corrected P-value threshold of 2.6×10^{-6}. Gene set P values were obtained using a competitive analysis that tests whether genes in a gene set are more strongly associated with the phenotype than other gene sets. We used European-ancestry subjects from the 1000 Genomes Project (Phase 3 v5a, MAF ≥ 0.01)[115] for the LD reference. The gene window used was 35 kb upstream and 10 kb downstream to include regulatory elements.

Gene sets were from two main sources. First, we included gene sets previously shown to be important for psychiatric disorders (71 gene sets; for example, FMRP binding partners, de novo mutations, GWAS top SNPs, ion channels)[57,116,117]. Second, we included gene sets from MSigDB (v5.2)[118], which includes canonical pathways and Gene Ontology gene sets. Canonical pathways were curated from BioCarta, KEGG, Matrisome, the Pathway Interaction Database, Reactome, Sigma-Aldrich, Signaling Gateway, Signal Transduction KE, and SuperArray. Pathways containing from 10–10,000 genes were included.

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

To evaluate gene sets related to antidepressants, gene sets were extracted from the Drug–Gene Interaction database (DGIdb v.2.0)[119] and the NIMH Psychoactive Drug Screening Program Ki DB[120] downloaded in June 2016. The association of 3,885 drug gene sets with major depression was estimated using MAGMA (v1.6). The drug gene sets were ordered by *P* value, and the Wilcoxon–Mann–Whitney test was used to assess whether the 42 antidepressant gene sets in the dataset (ATC code N06A in the Anatomical Therapeutic Chemical Classification System) had a higher ranking than expected by chance.

One issue is that some gene sets contain overlapping genes, and these may reflect largely overlapping results. The pathway map was constructed using the kernel generative topographic mapping algorithm (k-GTM) as described by Olier et al.[121]. GTM is a probabilistic alternative to Kohonen maps: the kernel variant is used when the input is a similarity matrix. The GTM and k-GTM algorithms are implemented in GTMapTool (see URLs). We used the Jaccard similarity matrix of FDR-significant pathways as input for the algorithm, where each pathway is encoded by a vector of binary values representing the presence (1) or absence (0) of a gene. The parameters for the k-GTM algorithm are the square root of the number of grid points (k), the square root of the number of RBF functions (m), the regularization coefficient (l), the RBF width factor (w), and the number of feature space dimensions for the kernel algorithm (b). We set k equal to the square root of the number of pathways, m equal to the square root of k, $l = 1$ (default), $w = 1$ (default), and b equal to the number of principal components explaining 99.5% of the variance in the kernel matrix. The output of the program is a set of coordinates representing the average positions of pathways on a 2D map. The x and y axes represent the dimensions of a 2D latent space. The pathway coordinates and corresponding MAGMA *P* values were used to build the pathway activity landscape using the kriging interpolation algorithm implemented in the R gstat package.

Mendelian randomization. We conducted bidirectional MR[122] analysis for four traits: years of education (EDY)[62], body mass index (BMI)[29], coronary artery disease (CAD)[63], and schizophrenia (SCZ)[19]. We denote z as a genetic variant (a SNP) that is significantly associated with x, an exposure or putative causal trait for y (the disease/trait outcome). The effect size of x on y can be estimated using a two-step least-squares (2SLS)[123] approach: $\hat{b}_{xy} = \hat{b}_{zy}/\hat{b}_{zx}$, where $\hat{b}_{zx}$ is the estimated effect size for the SNP–trait association for the exposure trait and $\hat{b}_{zy}$ is the effect size estimated for the same SNP in the GWAS of the outcome trait.

We used generalized summary statistics–based MR (GSMR)[64] to estimate $\hat{b}_{xy}$ and its standard error from multiple SNPs associated with the exposure trait at a genome-wide significance level. We conducted bidirectional GSMR analyses for each pair of traits and report results after excluding SNPs that failed the HEIDI outlier heterogeneity test (which is more conservative than excluding SNPs that have an outlying association likely driven by locus-specific pleiotropy). GSMR is more powerful than inverse-weighted MR (IVW-MR) and MR-Egger because it takes account of the sampling variation of both $\hat{b}_{zx}$ and $\hat{b}_{zy}$. GSMR also accounts for residual LD between the clumped SNPs. For comparison, we also conducted IVW-MR and MR-Egger analyses[124]. More details are provided in the Supplementary Note.

Genome build. All genomic coordinates are given in NCBI Build 37/UCSC hg19.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The PGC's policy is to make genome-wide summary results public. Summary statistics for a combined meta-analysis of PGC29 with five of the six expanded samples (deCODE, Generation Scotland, GERA, iPSYCH, and UK Biobank) are available on the PGC website (see URLs). Results for 10,000 SNPs for all seven cohorts are also available on the PGC website.

Genome-wide association summary statistics for the Hyde et al. cohort (23andMe, Inc.) must be obtained separately. These can be obtained by qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Contact David Hinds (dhinds@23andme.com) to apply for access to the data. Researchers who have the 23andMe summary statistics can readily recreate our results by performing meta-analysis of the six-cohort results file with the Hyde et al. results file from 23andMe[28].

The availability of genotype data for PGC29 is described in Supplementary Table 15. For the expanded cohorts, interested users should contact the lead principal investigators of these cohorts (which are separate from the PGC).

References

97. World Health Organization. *International Classification of Diseases* (World Health Organization, Geneva, 1978).
98. World Health Organization. *International Classification of Diseases* (World Health Organization, Geneva, 1992).
99. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, Washington, DC, 1994).
100. Abecasis, G. R. et al. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
101. ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
102. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
103. Vernot, B. et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235–239 (2016).
104. Bryois, J. et al. Evaluation of chromatin accessibility in prefrontal cortex of schizophrenia cases and controls. *Nat. Commun.* (in the press).
105. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
106. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
107. Finucane, H. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* https://dx.doi.org/10.1038/s41588-018-0081-4 (2018).
108. Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017).
109. Jansen, R. et al. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum. Mol. Genet.* **26**, 1444–1451 (2017).
110. Lappalainen, T. et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
111. Ng, B. et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).
112. Hannon, E., Weedon, M., Bray, N., O'Donovan, M. & Mill, J. Pleiotropic effects of trait-associated genetic variation on DNA methylation: utility for refining GWAS loci. *Am. J. Hum. Genet.* **100**, 954–959 (2017).
113. de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat. Rev. Genet.* **17**, 353–364 (2016).
114. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
115. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
116. Turner, T. N. et al. denovo-db: a compendium of human de novo variants. *Nucleic Acids Res.* **45** (D1), D804–D811 (2017).
117. Pirooznia, M. et al. High-throughput sequencing of the synaptome in major depressive disorder. *Mol. Psychiatry* **21**, 650–655 (2016).
118. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
119. Wagner, A. H. et al. DGIdb 2.0: mining clinically relevant drug–gene interactions. *Nucleic Acids Res.* **44** (D1), D1036–D1044 (2016).
120. Roth, B. L., Kroeze, W. K., Patel, S. & Lopez, E. The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrasment of riches? *Neuroscientist* **6**, 252–262 (2000).
121. Olier, I., Vellido, A. & Giraldo, J. Kernel generative topographic mapping. in *ESANN 2010 Proc.* 28–30 (2010).
122. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
123. Wooldridge, J. M. *Introductory Econometrics: A Modern Approach* (Cengage Learning, Boston, MA, 2015).
124. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

nature research

Corresponding author(s): Naomi Wray and Patrick Sullivan

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	This is fully described in the Online Methods and associated Supplemental Tables
Data analysis	Software used is described in the Online Methods.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

See the last two paragraphs in the Online Methods and Table S14.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Variable by analysis but given in the manuscript
Data exclusions	Exclusions done during GWAS quality control; described in detail in the paper
Replication	NA
Randomization	NA
Blinding	This was not practical given our design.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

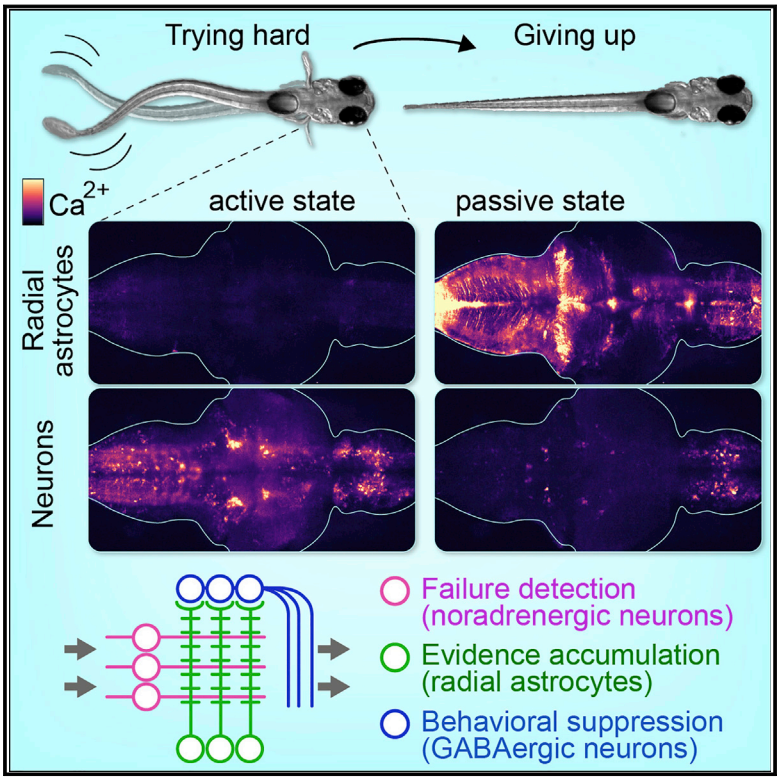
Population characteristics	This is fully described in the Online Methods and associated Supplemental Tables
Recruitment	This is fully described in the Online Methods and associated Supplemental Tables

Article

Cell

Glia Accumulate Evidence that Actions Are Futile and Suppress Unsuccessful Behavior

Graphical Abstract



Highlights

- Zebrafish "give up" after swim attempts repeatedly fail to generate movement
- Whole-brain imaging reveals noradrenergic neurons that encode swim failures
- Radial astrocytes accumulate the noradrenergic failure signal and trigger passivity
- Glia accumulate evidence that actions are futile, then switch behavioral state

Authors

Yu Mu, Davis V. Bennett, Mikail Rubinov, ..., Brett D. Mensh, Loren L. Looger, Misha B. Ahrens

Correspondence

muy@janelia.hhmi.org (Y.M.), ahrensm@janelia.hhmi.org (M.B.A.)

In Brief

Whole-brain imaging in virtual-reality-immersed zebrafish reveals that failed swim attempts are detected by noradrenergic neurons, which drive glial cells that accumulate calcium until they trigger the suppression of further futile attempts.

Mu et al., 2019, Cell *178*, 1–17
June 27, 2019 © 2019 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2019.05.050

CellPress

Please cite this article in press as: Mu et al., Glia Accumulate Evidence that Actions Are Futile and Suppress Unsuccessful Behavior, Cell (2019), https://doi.org/10.1016/j.cell.2019.05.050

Article

Cell

Glia Accumulate Evidence that Actions Are Futile and Suppress Unsuccessful Behavior

Yu Mu,[1,4,*] Davis V. Bennett,[1,2,4] Mikail Rubinov,[1,3,4] Sujatha Narayan,[1] Chao-Tsung Yang,[1] Masashi Tanimoto,[1] Brett D. Mensh,[1] Loren L. Looger,[1] and Misha B. Ahrens[1,5,*]

[1]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA
[2]Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL, USA
[3]Department of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: muy@janelia.hhmi.org (Y.M.), ahrensm@janelia.hhmi.org (M.B.A.)
https://doi.org/10.1016/j.cell.2019.05.050

SUMMARY

When a behavior repeatedly fails to achieve its goal, animals often give up and become passive, which can be strategic for preserving energy or regrouping between attempts. It is unknown how the brain identifies behavioral failures and mediates this behavioral-state switch. In larval zebrafish swimming in virtual reality, visual feedback can be withheld so that swim attempts fail to trigger expected visual flow. After tens of seconds of such motor futility, animals became passive for similar durations. Whole-brain calcium imaging revealed noradrenergic neurons that responded specifically to failed swim attempts and radial astrocytes whose calcium levels accumulated with increasing numbers of failed attempts. Using cell ablation and optogenetic or chemogenetic activation, we found that noradrenergic neurons progressively activated brainstem radial astrocytes, which then suppressed swimming. Thus, radial astrocytes perform a computation critical for behavior: they accumulate evidence that current actions are ineffective and consequently drive changes in behavioral states.

INTRODUCTION

Perseverance can lead to success, but some challenges are best met by high-activity bursts separated by behavioral quiescence. This pattern is often seen in motile animals that find themselves unable to move. Captured prey may become passive while waiting for a predator to loosen its grip before trying hard to escape; a fish stuck in mud may conserve energy between vigorous attempts to free itself. Periods of passivity are also observed during laboratory tasks in which an animal's actions fail to overcome challenges (e.g., passive coping, learned helplessness) (Andalman et al., 2019; Maier, 1984; Warden et al., 2012) or fail to acquire rewards (Salamone et al., 2016). Here, we studied the mechanisms of a form of "futility-induced passivity" in zebrafish to ask: how does the brain detect failures, accumulate evidence over time to infer that its motor patterns (attempts) are futile, and switch behavioral strategy?

To search for neurons involved in behavioral-state switching, we used whole-brain light-sheet imaging to survey single-cell dynamics during behavior in larval zebrafish. Because astrocytes influence neuronal activity (Bazargani and Attwell, 2016; Nedergaard, 1994; Parpura et al., 1994), we also imaged a glial cell type in zebrafish we refer to as radial astrocytes (a cell type previously also referred to in zebrafish as radial astroglia, astroglia, radial glia) (Cuoghi and Mola, 2009; Grupp et al., 2010). These cells have an apical-basal projection and bushy processes, similar to mammalian radial glia and astrocytes, respectively (Lyons and Talbot, 2014).

Interactions between astrocytes and neurons have been investigated extensively. Astrocytes can respond to neuronal activity by intracellular calcium increase. They can, in turn, influence other neurons through multiple mechanisms (Bazargani and Attwell, 2016), such as transmitter release (Araque et al., 2014) or modulation of extracellular potassium concentration (Haydon and Nedergaard, 2014; Wang et al., 2015). They play a role in controlling the states of neural networks, perform spatial and temporal integration, and can underlie slow shifts in brain states (Araque et al., 2014) like sleep and seizures (Kjaerby et al., 2017; Verdugo et al., 2019). Calcium signals in astrocytes are triggered by sensory input and behavior (Bekar et al., 2008; Ma et al., 2016; Paukert et al., 2014; Schummers et al., 2008; Srinivasan et al., 2015; Verkhratsky et al., 1998). Astrocytes can modulate circuit function and behavior (Cui et al., 2018; Gourine et al., 2010; Morquette et al., 2015; Yu et al., 2018). They can enhance memories (Adamsky et al., 2018), integrate neuronal signals over time (Deemyad et al., 2018), respond to neuromodulators including norepinephrine (NE) (Bekar et al., 2008; Ma et al., 2016; Paukert et al., 2014; Srinivasan et al., 2015; Verkhratsky et al., 1998), and serve as intermediaries between neuromodulators and neurons (Ma et al., 2016; Pabst et al., 2016). These findings raise the possibility that astrocytes play further roles typically ascribed to neurons, such as driving rapid behavioral-state switches.

To investigate brain-wide neuronal and astrocytic activity during futility-induced passivity, we used a virtual-reality (VR) environment to first simulate realistic visual feedback during attempted swimming; we then suddenly withheld this feedback to

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

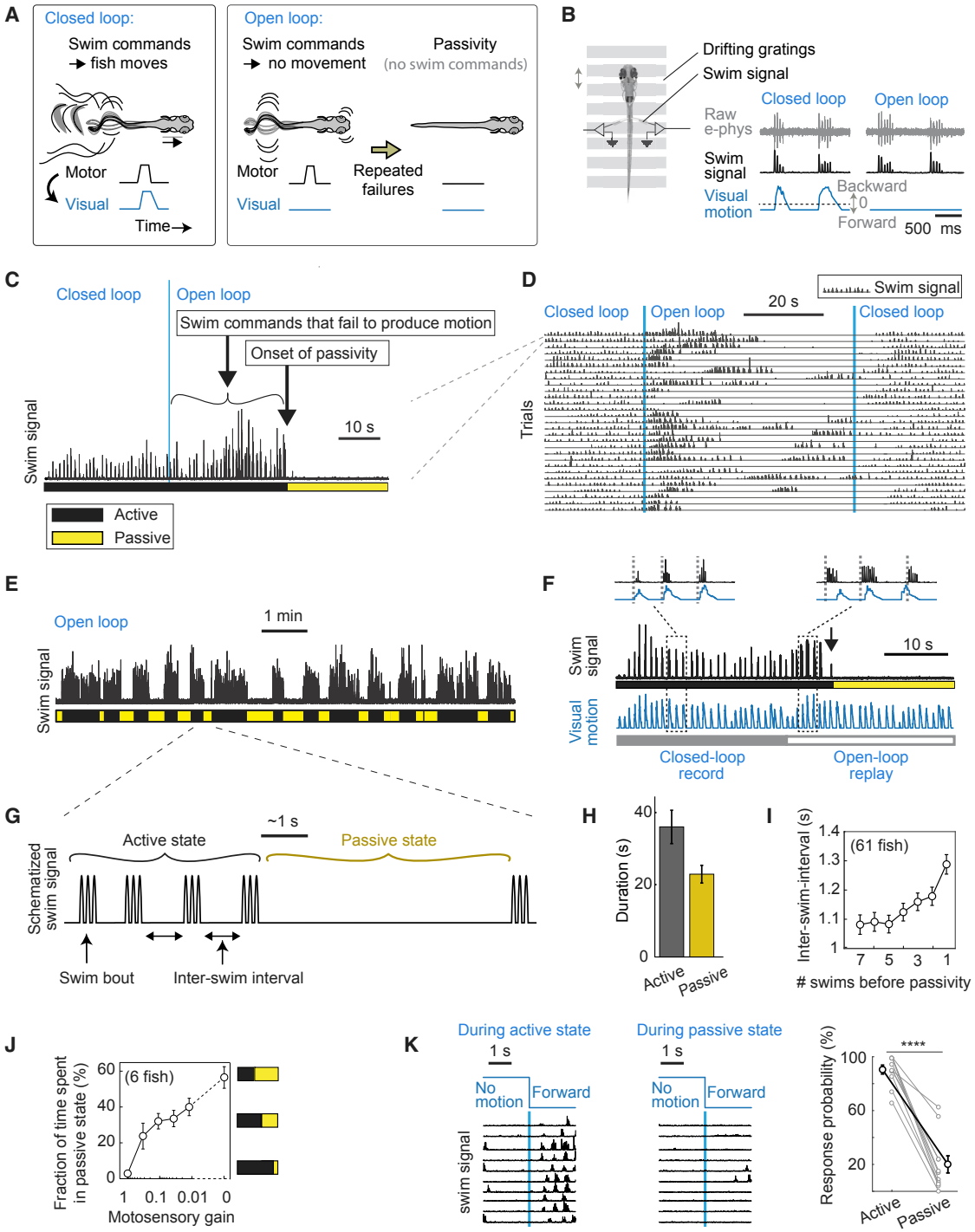


Figure 1. Lack of Visual Feedback during Swimming Triggers Passive Behavioral State

Figure360▷ For a Figure360 author presentation of this figure, see https://doi.org/10.1016/j.cell.2019.05.050.

(A) Locomotion in closed loop versus open loop. After repeated swim failures, zebrafish entered a passive behavioral state.

(B) Virtual environment for fictively swimming paralyzed fish. Unless indicated otherwise, all VR experiments used a constant forward-drifting grating to induce motor behavior. In closed loop, the visual scene was accelerated backward as fish attempted to swim; swim commands were detected from motor neuron axons with tail-attached electrodes. In open loop, the visual scene did not respond to swim commands. Dashed line, stationary grating; above dashed line, backward motion; Below, forward motion.

(C) Futility-induced passivity. In closed loop, fish swam naturalistically, at about one bout per second. In open loop, fish first swam with higher vigor, then stopped swimming for several to tens of seconds.

(legend continued on next page)

Please cite this article in press as: Mu et al., Glia Accumulate Evidence that Actions Are Futile and Suppress Unsuccessful Behavior, Cell (2019), https://doi.org/10.1016/j.cell.2019.05.050

render swim efforts ineffective. The animals responded by first increasing their motor vigor, then suddenly stopping swimming, followed by cycling between passive and vigorously active behavioral states. To mediate such behavior, the brain must detect movement failures (when a motor command does not lead to perceived motion), integrate the failure signal over multiple attempts (to infer that the attempts are futile), and switch to another behavioral strategy. Using whole-brain imaging and targeted perturbations, we found that the noradrenergic (NE) system represents sensory-motor mismatch during movement failures, that radial astrocytes accumulate these NE mismatch signals, and that resulting surges in astrocytic activity activate downstream neurons and trigger behavioral passivity. These results show that radial astrocytes are fundamental to the rapid regulation of behavioral states.

RESULTS

Withholding Visual Feedback during Swimming Induces Passivity

Since swimming is central to fish behavior, we focused on the animal's response when it perceives that swimming is futile. We created this perception by withholding the visual feedback that normally accompanies motor output in a fictive swimming VR environment. After repeatedly attempting to swim with no change in visual flow, animals became passive (Figure 1A).

Animals (5–7 days post fertilization) were placed in a virtual environment where they fictively swam in response to forward-drifting visual scenes, as if against a water current, during which they attempted to stay in the approximate same spatial location (Orger et al., 2000; Rock and Smith, 1986; Vladimirov et al., 2014). In this "closed-loop" environment (Figure 1B), swim commands, measured from motor neuron firing in the tail, triggered backward visual flow like that seen during swim bouts. This elicited naturalistic swimming at a regular rate of about one swim bout per second (Dunn et al., 2016; Kawashima et al., 2016). The fish closely followed the visual stimulus, adjusting their swim vigor to compensate for changes in motosensory gain (i.e., coupling efficiency between swimming and visual feedback) (Ahrens et al., 2012; Portugues and Engert, 2011).

Fish were then switched to an "open-loop" protocol where forward visual flow was seen but swimming did not elicit visual feedback (Figures 1A and 1B). The animals increased swim vigor for many swim bouts but then abruptly became passive for a long time (Figures 1C and 1D), after which they resumed swimming, cycling between passive and active behavioral states lasting several to tens of seconds (Figure 1E). The animals also switched to a passive state when viewing replayed visual motion patterns in open loop that had been previously recorded during closed loop (Figure 1F), indicating that the passivity resulted from decoupling sensory feedback and behavior. Passive states (epochs when the animal did not swim for at least 5 s; Figures S1A and S1B) lasted 24.4 ± 2.5 s (mean ± SEM), and active states lasted 36.1 ± 4.2 s (n = 74; Figure 1H).

Several seconds prior to passivity, fish swam more vigorously and less frequently and attempted to make larger amplitude turns (Figures 1I and S1C). They were also passive when there was some, but very little, visual feedback (very low motosensory gain; Figure 1J).

Passivity was not due to fatigue, as similar swim vigor was accompanied by different passivity durations in the lowest gain range (Figures S1E and S1F). Nor was passivity due to struggles, as most passive periods were not preceded by struggling (Figures S1G–S1I). Thus, the crucial driver of the switch to passive behavior is the absence of visual feedback during swimming.

During passivity, fish were almost non-responsive to moving visual stimuli (Figures 1K and S1J), yet neuronal activity increased in the abducens nucleus, which controls eye movements (Figure S1K), suggesting that fish attempted to move their eyes more during this time.

The alternation of two opposing behavioral states allowed us to study the mechanisms of brain-state transitions.

Correlates of Passivity in Neuronal and Astrocytic Activity

We hypothesized that the passive state might be triggered by increased activity in cells that suppress swimming. We performed whole-brain cell-resolution calcium imaging in neurons and radial astrocytes during futility-induced passivity (Figure 2A). We used a light-sheet microscope (Ahrens et al., 2013; Panier et al., 2013; Vladimirov et al., 2014) with the VR environment to image animals expressing either GCaMP6f (Chen et al., 2013) or jRGECO1b (Dana et al., 2016) in almost all neurons under the *elavl3* promoter (Park et al., 2000) or in radial astrocytes under the *gfap* promoter (Bernardos and Raymond, 2006). We imaged either cell type or both with simultaneous dual-color imaging (STAR Methods).

When animals switched from the active to the passive behavioral state, whole-brain average neuronal activity decreased,

(D) Multiple trials of futility-induced passivity from a typical fish, showing passivity and switches to active behavior. Trials are successive epochs from a continuous recording, aligned by the onset of open loop.

(E) In prolonged open loop (tens of minutes), fish cycled between active and passive behavioral states.

(F) Futility-induced passivity also occurred when the stimulus seen in closed loop was replayed in open loop, showing that passivity was not caused by a change in the pattern of the visual motion. Dashed vertical lines indicate alignment of motor and sensory events on left and misalignment on right.

(G) Diagram with terminology for behavioral transitions.

(H) Mean duration (time between behavioral transitions) of active and passive epochs in prolonged open loop, n = 74 fish. An epoch is passive if the fish did not swim for 5 s or more (see Figures S1A and S1B). Error bars, SEM in all figures.

(I) Before the switch to passivity, inter-swim interval progressively increased, n = 61 fish (see Figure S1E).

(J) Prevalence of the passive behavioral state depended on the motosensory gain (coupling efficiency between swimming and visual feedback). Naturalistic gain (≈1) induced an active state. As the gain approached 0 (i.e., open loop), fish gradually spent more and more time in the passive state. n = 6 fish.

(K) Passive fish were less responsive to visual stimuli. Active or passive states were elicited via closed loop or open loop. Then a brief static stimulus was followed by brief forward motion. In passive fish, response to motion was reduced. Left: example. Right: population data. Two-tailed paired t test, $p = 4.2 \times 10^{-7}$, n = 11 fish (see Figure S1J).

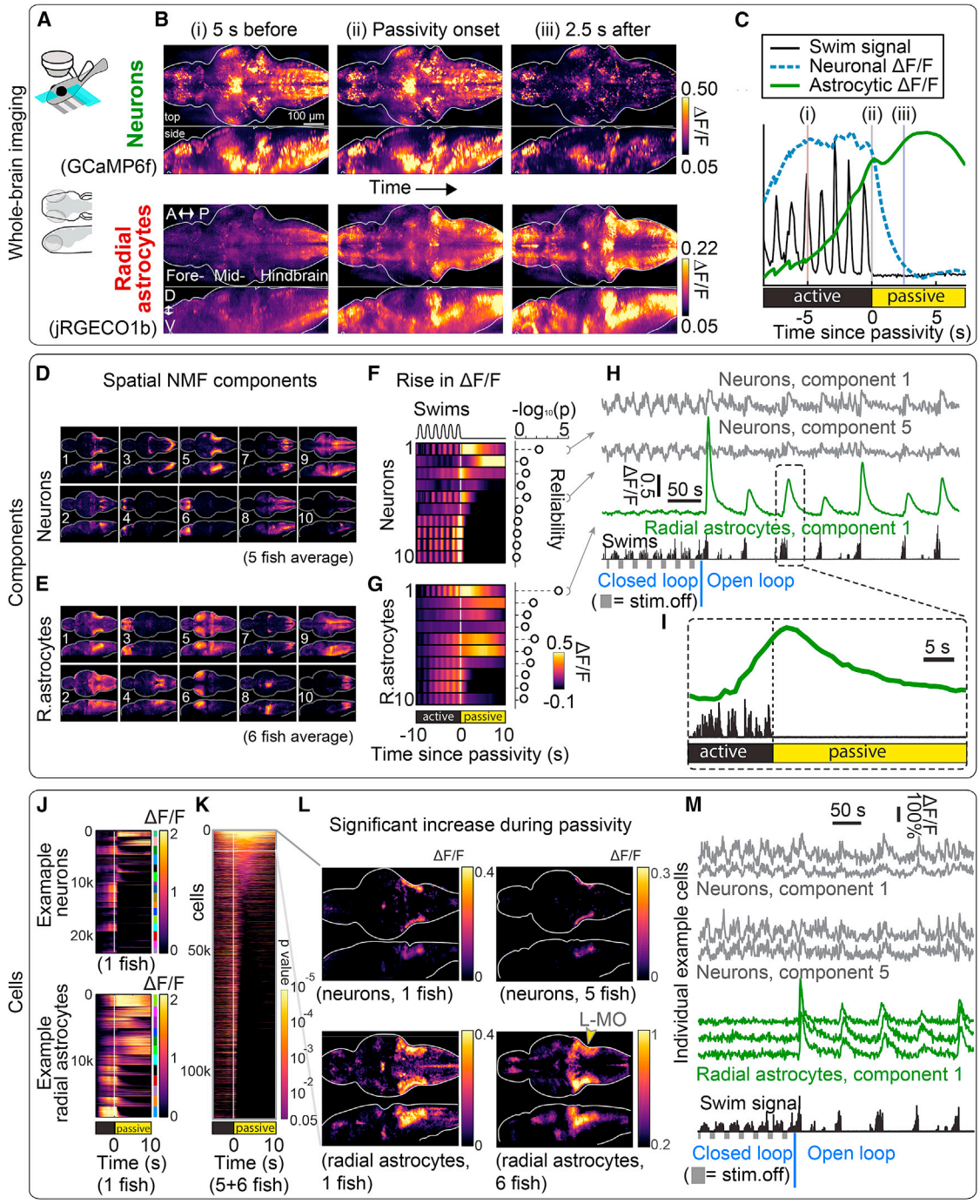


Figure 2. Whole-Brain Neuronal and Astrocyte Dynamics Show Ca^{2+} Increases in Specific Cell Populations during Futility-Induced Passivity

Figure360▷ For a Figure360 author presentation of this figure, see https://doi.org/10.1016/j.cell.2019.05.050.

(A) Experimental setup. A light-sheet microscope recorded from most radial astrocytes and neurons in the brain at cellular resolution, using either single-population GCaMP6f or simultaneous dual-color GCaMP6f and jRGECO1b recordings, while fictive behavior was monitored and visual stimulus was delivered.

(B) Whole-brain neuronal and glial activity before, at, and after passivity onset in an example *Tg(gfap:jRGECO1b); Tg(elavl3:GCaMP6f)* fish. Top: mean neuronal brain-wide activity was higher during active behavior than passivity. Bottom: conversely, mean brain-wide glial Ca^{2+} signals increased before passivity onset and remained elevated during passivity. A, anterior; P, posterior; D, dorsal; V, ventral. Images are maximum-intensity projections of means triggered on passivity onset (64 events).

(C) Neuronal and glial signals near passivity onset, averaged across the brain of the same fish from (B). Average neuronal Ca^{2+} decayed after passivity onset, while average glial Ca^{2+} increased before passivity onset and peaked soon after.

(legend continued on next page)

while glial activity increased (Figure 2B; Video S1). This increase in glial Ca^{2+} started seconds before passivity onset (Figures 2B and 2C). Both neuronal and glial Ca^{2+}, averaged across the brain, were higher and fluctuated more strongly in open loop than in closed loop (Figure S2A). Fluctuations in glial Ca^{2+} appeared strongest and most widespread during the first several switches to passivity in the minutes following the start of open loop, then became weaker and more localized.

To quantify these activity patterns in more detail and identify cell populations with similar activity patterns, we used non-negative matrix factorization (NMF) in neurons and radial astrocytes separately. A first round of NMF extracted single neurons, radial astrocytes, and segments of processes based on co-fluctuating activity of neighboring voxels (called "cell segments," or "cells" for brevity; STAR Methods). A second round factorized the brain into spatial components—sets of cells with coordinated activity (Figures 2D and 2E; Video S2)—and temporal components that describe the activity patterns of these cell groups (Figures 2F and 2G). We clustered the components into a set that occurred consistently across all fish (STAR Methods). Multiple components showed activity increases near the onset of passivity, including neuronal and glial components with elevated activity in the passive period (Figures 2F and 2G). A response reliability measure ($-\log_{10}$ p; STAR Methods) that quantifies consistency over repeated behavioral switches identified a consistent subset of glial signals (Figure 2G, component 1). Calcium in astrocyte component 1 increased before and peaked after passivity onset (Figures 2G–2I). When fish in a closed-loop protocol stopped swimming because the stimulus stopped moving, glial Ca^{2+} signals were weaker than during futility-induced passivity (Figure 2H). The most consistent surge in glial Ca^{2+} occurred in a neuropil-rich area in the lateral hindbrain, which we call area L-MO (lateral medulla oblongata, spanning rhombomeres 2–6, the area of astrocyte component 1 in Figure 2E; also called "neuropil area 4" in the Z-Brain atlas [Randlett et al., 2015] and regions #12 and #171 in Zebrafish Brain Browser [Marquart et al., 2015]).

Finally, we analyzed single-cell segments. Ca^{2+} signals from neurons and glia were analyzed near passivity onset (Figure 2J). For each cell, we used a p value to quantify the consistency of Ca^{2+} increases above baseline (Figure 2K; Video S3; STAR Methods). Of all brain regions, L-MO radial astrocytes showed the most consistent increases in activity during passivity (Figure 2L) and showed ramping Ca^{2+} signals before passivity onset (Figures 2G, 2I, 2M, and S2C–S2E). Similar glial Ca^{2+} activity occurred in another futility-induced passivity assay, where untethered fish become passive after being submerged in a viscous agarose solution (Figure S2F; Video S4).

Astrocytic Ca^{2+} started to increase before the switch to passivity (Figures 2G, 2I, and S2C–S2E). This Ca^{2+} elevation and the peak that followed at passivity onset led us to hypothesize that radial astrocytes may suppress swimming.

Anatomy of Radial Astrocytes

To provide anatomical context for these observations, we examined the morphology of radial astrocytes extending processes to the L-MO. In fish expressing GCaMP6f in radial astrocytes, some residual GCaMP6f may remain in newly differentiated neurons (Anthony et al., 2004; Fujita et al., 2014; Sloan and Barres, 2014). We examined protein expression driven by *gfap* and *elavl3* promoters and found them well segregated across the brain (Figures S3A and S3B), suggesting, together with the differences in Ca^{2+} activity (Figures 2F and 2G), that neuronal and glial Ca^{2+} measurements were well separated.

In *gfap* transgenics (Figures 3A–3C), we filled single radial astrocytes with dye to visualize their morphology (Figure 3D). The cell bodies resided in the dorsal midline of the hindbrain in rhombomeres 2–6, and the bushy processes resided in L-MO; the Ca^{2+} signals locked to passivity originate in these processes (Figures 3 and S3C–S3F). These processes ramify throughout the L-MO neuropil and appose GABAergic cell bodies at the edge of the brain (Figure 3D). The GCaMP6f signal in the L-MO can spread to the cell bodies (Figure 2D, radial astrocyte components 1→2→4; Video S1) through the connecting processes. Activation of other brain areas (e.g., glial components 3, 5, 6) happened more often right after transitions from a closed-loop to an open-loop protocol. Radial astrocytes lacked active membrane properties and had a more negative resting-membrane potential than neurons (Figure S3G), consistent with previous observations of these cells (Li et al., 2015).

(D and E) Spatial NMF components. Neuronal and glial data recorded with GCaMP6f in separate transgenic lines. Brightness: spatial NMF weights. n = 5 fish (neurons), n = 6 fish (radial astrocytes).

(F and G) Temporal neuronal and glial Ca^{2+} signals, per NMF component of (D) and (E), averaged over fish and trials, centered on passivity onset. Right: reliability ($-\log_{10}$(p)), calculated by comparing Ca^{2+} activity 1–4 s after passivity onset to mean activity 10–5 s before passivity onset (higher reliability: more consistent responses over trials). Multiple neuronal and glial components had strong correlations to behavior. Neuronal and glial component 1 overlaps with L-MO.

(H) Signals from example fish; see also Figure S2B. Fish was first in closed loop, presented with alternating moving and stationary gratings (fish swam little during stationary gratings), and next in open loop, wherein animals cycled through active and passive states. Radial astrocytes were engaged in switches to passivity in open loop, not swim pauses in closed loop.

(I) Astrocytic signal from (H), showing a rise before and peak after passivity onset.

(J) Cellular signals from segmented neurons and glia from two example fish, grouped by component (colored bars, right), relative to passivity onset.

(K) Neurons and glia (random subset) from all fish with elevated signals after passivity onset, ordered by p value from a statistical test of how reliably Ca^{2+} is increased above baseline (10–5 s before passivity).

(L) Brain maps of neurons and radial astrocytes with reliably elevated Ca^{2+} just after passivity onset (1–4 s) relative to baseline. L-MO signals were most consistently above baseline across fish. Left: example. Right: population average (5 fish for neurons, 6 for glia, registered to standard brain). p values were not used to test statistical significance but to create an exploratory functional brain map (STAR Methods).

(M) Ca^{2+} traces from single cells, same fish as in (H) and (I), during closed loop (left) and open loop (right). Radial astrocytes selectively respond to futility-induced passivity.

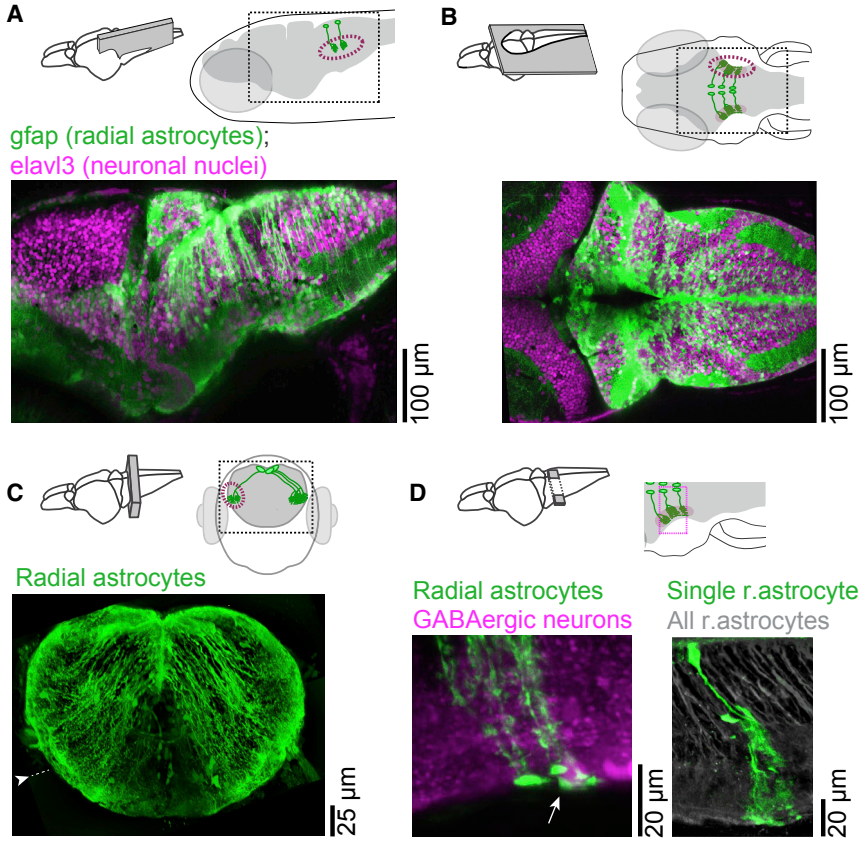


Figure 3. Morphology of Radial Astrocytes Projecting to L-MO

(A) Maximum intensity projection (MIP) along the medial-lateral axis of a 20 μm section from *Tg(elavl3:H2B-GCaMP6f);Tg(gfap:tdTomato)* fish. Dashed line, L-MO location. Neurons, magenta; glia, green.

(B) MIP along dorsal-ventral axis of a 20 μm section, same fish as (A). Radial astrocytes ramify in L-MO.

(C) MIP along the rostral-caudal axis of a 20 μm section in hindbrain. Radial astrocyte processes project ventral-laterally. Scale bar is corrected for 4× expansion (Freifeld et al., 2017; Tillberg et al., 2016) to approximate *in vivo* dimensions. Arrowhead indicates microscopy software stitching artifact.

(D) MIP along dorsal-ventral axis. Diagonal section with one (right) or several (left) radial astrocytes electroporated with synthetic dye. Astrocytic somata are proximal to dorsal midline of hindbrain; their processes run ventral-laterally, are dense in lateral hindbrain (green at brain border), and appose GABAergic somata (left: white arrow, magenta, labeled by *Tg(gad1b:RFP)*); radial astrocytes (right: gray) labeled by *Tg(gfap:jRGECO1b)*.

Radial Astrocyte Activation Increases Passivity; Silencing Decreases Passivity

Since radial astrocytes were activated before and at the transition from active to passive states, we asked whether disabling these cells prevents the behavioral-state transition and whether activating them triggers it. We used two-photon laser ablation to kill glial cells projecting to L-MO (Vladimirov et al., 2018) (~300 of ~1,000 such cells, using nuclear-localized GCaMP6f under the *gfap* promoter to target somata; STAR Methods) (Figure 4A) and assayed for futility-induced passivity. Ablated animals spent half as much time in the passive state as they did before ablation (Figures 4B and 4C), suggesting that radial astrocytes are necessary to trigger passivity. We confirmed that this effect was not restricted to fictive swimming: when deprived of visual feedback during tail motion, nonparalyzed head-embedded fish were more active after ablation (Figures 4D and S4A–S4C). This behavioral change was specific to radial astrocyte ablation, since ablation of either a matched number of neurons in random locations or a large part of the forebrain (STAR Methods) did not affect futility-induced passivity (Figures 4E, S4D, and S4E). This effect was not due to hyperactivity, since fish with glial ablation can still stop swimming and perform motor adaptation (Figures S4F and S4G). An additional nitroreductase-based chemogenetic approach to render radial astrocytes unhealthy (Tabor et al., 2014) also reduced passivity (Figure S4H). Thus, futility-induced passivity depends on the presence of healthy, functioning radial astrocytes.

Since astrocytes support neuron health (Banker, 1980; Lyons and Talbot, 2014), we controlled for adverse effects of glial ablation on neurons. After ablations, we observed no significant difference in swim frequency, power, or velocity in freely swimming fish (Figure 4F). Also, neural activity in L-MO was similar before and after glial ablation during closed-loop fictive swimming (Figure 4G), suggesting that the targeted ablation minimally affects neural activity outside of futility-induced passivity.

Finally, we blocked the inositol 1,4,5-trisphosphate receptor (IP_3R) to reduce Ca^{2+} release from intracellular stores in radial astrocytes. Ca^{2+} transients dropped (Figure S4I) when we locally puffed the IP_3R inhibitor xestospongin C (XeC) onto astrocytic cell bodies in the dorsal midline of the hindbrain (Figure 4H). We minimized exposure of neurons to XeC by injecting into a region containing mostly glia. This inactivation method also reduced futility-induced passivity (Figure 4H).

Together, these results strongly suggest that radial astrocytes are necessary for futility-induced passivity.

To test whether surges in astrocytic Ca^{2+} are sufficient to reduce swimming, we used a chemogenetic approach to increase intracellular Ca^{2+} in glia. We transiently expressed rat TRPV1 in a sparse subset (~1%) of radial astrocytes (Figure S5A; STAR Methods), rendering their membranes highly permeable to Ca^{2+} upon exposure to the agonist, capsaicin (other cells were unaffected since, unlike rat TRPV1, the zebrafish ortholog is insensitive to capsaicin) (Chen et al., 2016; Munns et al., 2015). We targeted only a tiny subset of cells to mitigate any unphysiological effects of TRPV1 activation. Application of capsaicin (5 μM) caused increased Ca^{2+} fluctuations in the population (Video S5), potentially through gap junctions (Fujii et al., 2017). The fish became passive more than 60% of the time, starting

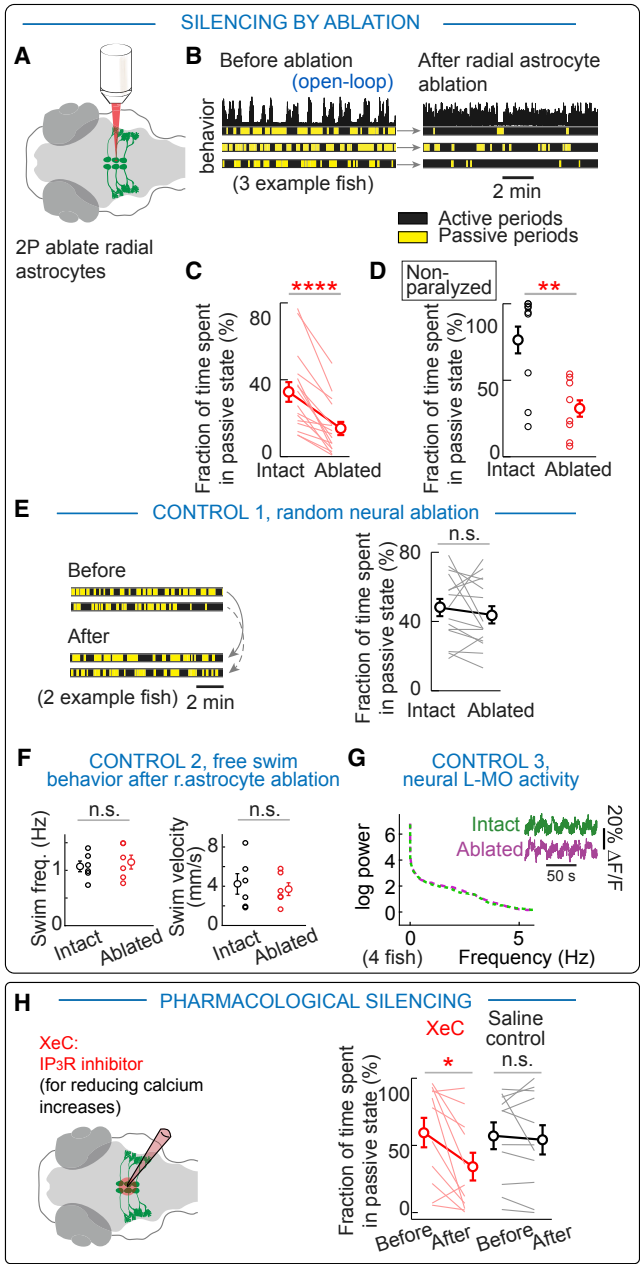


Figure 4. Radial Astrocytes Are Necessary for Futility-Induced Passivity

(A) Experimental design. ~300 astrocytes projecting to L-MO were ablated with a 2-photon laser with single-cell resolution, in fish with radial astrocytic nuclei fluorescently labeled. In all following panels, n.s., not significant; *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001.

(B) Reduced futility-induced passivity after glial ablation in 3 example fish.

(C) Population data showing a ~50% reduction in futility-induced passivity. Two-tailed paired t test, $p = 1.3 \times 10^{-5}$, n = 16.

(D) Alternative assay for futility-induced passivity: unparalyzed, head-restrained tail-free also show passivity. Fish with ablated astrocytes remained active most of the time (Figures S4A–S4C). Two-tailed non-paired t test, p = 0.0041, n = 8 fish in both conditions.

(E) Ablation of ~300 random neurons had no significant effect on passivity. Two-tailed paired t test, p = 0.34, n = 15 fish.

(F) Control for nonspecific effects of astrocytic ablation. During free swimming, ablated fish showed no abnormalities in swim frequency or velocity. Two-tailed non-paired t test, p = 0.75 for swim frequency, p = 0.67 for swim velocity, n = 6 fish.

(G) Imaging in L-MO before and after glia ablation showed no difference in neural activity in closed loop with alternating moving and stationary gratings. Inset, example activity; main panel, power spectral density. n = 4 fish.

(H) Reducing glial Ca^{2+} by blocking IP_3R reduced futility-induced passivity. XeC (100 μM inside pipette) or saline was puffed onto astrocytic somata. Two-tailed paired t test, p = 0.033, p = 0.54, n = 11 fish in both conditions.

when surges in astrocytic Ca^{2+} appeared (Figures 5A–5D) in a closed-loop environment with drifting gratings (where the normal passive time is 2.3% ± 1.2%). L-MO glial Ca^{2+} rose before passivity and peaked shortly thereafter (Figure 5E). This effect was absent when capsaicin was applied to fish without TRPV1 or saline to fish with TRPV1 (Figures 5F–5H). The temporal dynamics of capsaicin-triggered glial Ca^{2+} increases at passivity onset resembled those triggered by open loop in TRPV1-negative fish (Figure 5I), suggesting that TRPV1 evokes a naturalistic process. Unlike the switch to passivity induced by open loop, capsaicin-induced switching was not preceded by increased swim vigor (Figure S5B). Thus, surges in Ca^{2+} in L-MO radial astrocytes alone appear sufficient to stop swimming.

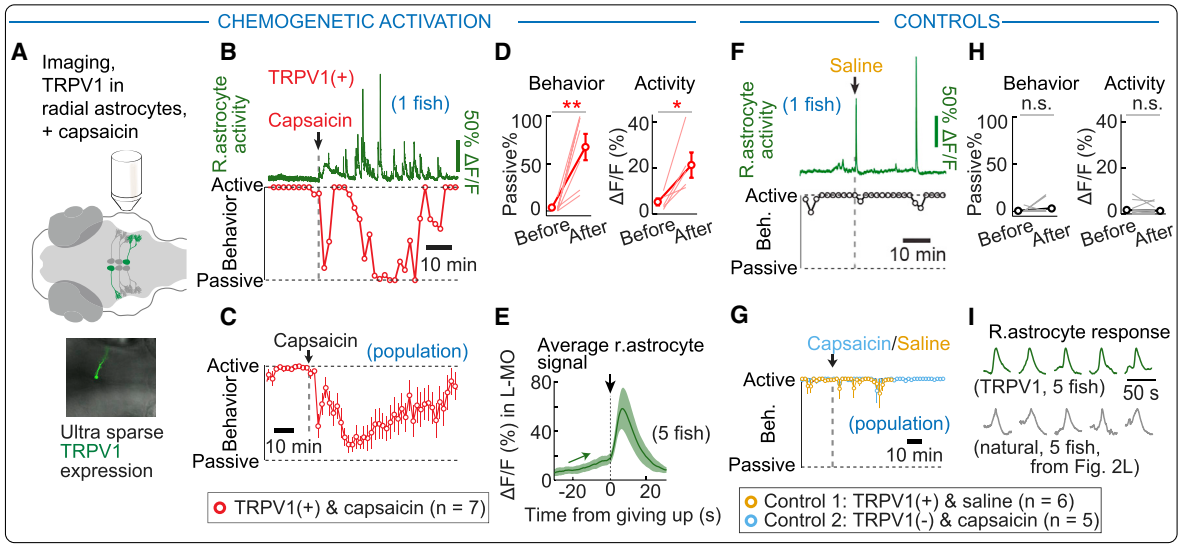
To test with temporal precision whether activating radial astrocytes suppresses swimming, we expressed CoChR (Klapoetke et al., 2014), a large-current channelrhodopsin, in radial astrocytes and activated it by shining blue light on L-MO with a digital micromirror device (DMD) (Zhu et al., 2012a). The resulting increase in glial Ca^{2+} (Figures 5J and S5C; Video S5) looked natural, including in its wave-like propagation (Figure S5C). However, opening this channel might elicit concomitant changes in intracellular sodium levels and pH, and although we did not detect double-labeled cells, neurons may express residual low levels of CoChR (Anthony et al., 2004; Fujita et al., 2014; Sloan and Barres, 2014). Hence, in another experiment, we expressed the optogenetic G protein-coupled receptor (GPCR) Opto-α1-AR (Airan et al., 2009) in radial astrocytes to mimic the Gq-coupled human α_{1a}-adrenergic receptor; in this case, exposure to blue light increased glial Ca^{2+}. Even if there were residual expression in neurons, this manipulation was unlikely to drive spikes due to relatively low efficiency and the indirect relationship between GPCR signaling and spiking. Both CoChR and Opto-α1-AR (Figure 5J) activation (Figueiredo et al., 2014) induced a drop in swim power (beyond a first mild drop also in controls, perhaps as the fish sees scattered blue light) (Figures 5K–5P, S5D, and S5E).

Thus we showed, with six different gain- and loss-of-function experiments, that activation of radial astrocytes can suppress swimming (Figure 5Q; caveats and controls in Table S1) and trigger the passive behavioral state.

Radial Astrocytes Drive Activity in GABAergic Neurons that Inhibit Swimming

We reasoned that any effect of an elevation in astrocytic Ca^{2+} on behavior must occur via neurons. Neurons at the edge of L-MO are reliably activated at passivity onset (Figures 2D, neuronal component 1, and S5F), suggesting that their activity increase is due to astrocytic Ca^{2+} increases. Supporting this

Please cite this article in press as: Mu et al., Glia Accumulate Evidence that Actions Are Futile and Suppress Unsuccessful Behavior, Cell (2019), https://doi.org/10.1016/j.cell.2019.05.050



(legend on next page)

idea, increasing astrocytic Ca^{2+} by either CoChR or TRPV1 activates neurons at the edge of L-MO (Figures 5R, S5G, and S5H). Optogenetic activation of these neurons, which are GABAergic with processes innervating L-MO (Figure 5R), also suppressed swimming (Figure 5S). These results provide evidence that L-MO radial astrocytes activate GABAergic neurons at the edge of L-MO and suggest that these cells in turn suppress swimming.

The Noradrenergic System Conveys Swim Failures to Radial Astrocytes

Futility-induced passivity results from multiple motor failures, which must be communicated to radial astrocytes by neurons. We hypothesized that the NE system might drive radial astrocytic Ca^{2+} signals based on known relationships of norepinephrine to behavior (Sara, 2009) and its effect on astrocytes (Bekar et al., 2008; Ding et al., 2013; Ma et al., 2016; Paukert et al., 2014; Slezak et al., 2018). We found that neurons in the locus coeruleus (LC) and the NE cluster of the medulla oblongata (NE-MO; Figures 6A and S6A) (Farrar et al., 2018; Tay et al., 2011) increased activity before switches to passivity. We imaged Ca^{2+} in NE neurons and in glia simultaneously in *Tg(th-P2A-Gal4); Tg(UAS:GCaMP6s); Tg(gfap:jRGECO1b)* fish. During futility-induced passivity, signals from NE neurons and glia were temporally coupled, with NE cell responses preceding astrocytic responses (Figures 6B and 6C; Video S6).

Radial astrocytes express multiple NE receptors (Table S2; Figure S6B), and NE axons project to L-MO (Figure S6C; Video S6), indicating that glial processes in L-MO could be driven by NE neurons.

Ablating NE-MO with a two-photon laser greatly reduced time spent in passivity, but no significant reduction resulted from ablation of the LC; ablation of the serotonergic dorsal raphe nucleus (DRN) also did not result in a significant change (Figures 6D–6F). Thus, the NE-MO, which may be homologous to NE cluster A2 in mammals (Rinaman, 2011), is necessary for futility-induced passivity.

To directly test whether NE-MO activates radial astrocytes and triggers passivity, we optogenetically activated NE-MO while imaging radial astrocytes (Figure 6G) (Zhu et al., 2012a) in fish expressing CoChR in NE cells and jRGECO1b in radial astrocytes (*Tg(dbh:KalTA4); Tg(UAS:CoChR-eGFP); Tg(gfap:jRGECO1b)*). This led to Ca^{2+} responses in glial processes in L-MO and induced passivity for ~10 s (Figures 6H and 6I). In fish with radial astrocytes ablated (as in Figures 4A–4C), activation did not lead to passivity (Figures 6I and S6E), showing that passivity resulting from artificial NE-MO activation also requires radial astrocytes.

We then identified the NE receptor that activates radial astrocytes (RNA sequencing, Table S2). After adding prazosin (α1-adrenergic receptor antagonist, 50 μM) or cyclazosin (α1B receptor antagonist, 10 μM), optogenetic activation of NE-MO no longer caused radial astrocyte activation (Figures 6J–6L). Puffing an α1 agonist elicited Ca^{2+} responses, and blocking the α2 receptor did not eliminate Ca^{2+} signals (Figures S6F and S6G). Thus, the response of L-MO radial astrocytes to NE is mediated by the α1B adrenergic receptor.

Finally, we tested whether astrocytic Ca^{2+} responses were driven solely by NE or by a combination of NE and local circuit activity like mammalian astrocytes in visual cortex (Paukert et al., 2014). We used swimming, which correlates with L-MO neural activity, as a proxy for local circuit activity (Figure 6M). Activating NE-MO during swimming caused a greater glial response than activating NE-MO without swimming (Figure 6N). Thus, radial astrocyte Ca^{2+} integrates NE and local circuit activity.

Figure 5. Activation of Radial Astrocytes Reduces Swimming

(A) Experimental design. Rat TRPV1 was expressed in ~1% of radial astrocytes in *Tg(gfap:jRGECO1b)* fish. Fish were exposed to the TRPV1 agonist capsaicin.

(B) Example fish with increased Ca^{2+} transients in L-MO and concurrent increase in passivity after capsaicin treatment. The effect diminished after ~1 h, maybe due to TRPV1 desensitization and internalization (Sanz-Salvador et al., 2012), showing that fish remain healthy, as they can swim normally. Dashed line (passive) means the fish is passive 100% of the time; dashed line (active) means it is passive 0% of the time (both measured in 2 min bins).

(C) Population data, showing increased passivity after activation of TRPV1. n = 7 fish.

(D) Population data of L-MO astrocytic Ca^{2+} increase and passivity with TRPV1 activation. Two-tailed paired t test, behavior: p = 0.0029, n = 7 fish. $\Delta F/F$: p = 0.027, n = 5 fish.

(E) Astrocytic Ca^{2+} in L-MO near passivity onset. Average signal increased just before passivity onset and peaked just after, as in futility-induced passivity (Figures 2G–2I). Shading, SEM.

(F) Control, example. No reduction of swimming or increase in astrocytic Ca^{2+} when saline was added to fish expressing TRPV1.

(G) Control, population data. No change in behavior when saline was added to fish expressing TRPV1 (n = 6 fish) or capsaicin to fish without TRPV1 (n = 5 fish).

(H) Summary of controls. Experiments as in (F) and (G) were pooled. Two-tailed paired t test, behavior: p = 0.48, n = 11 fish. $\Delta F/F$: p = 0.51, n = 7 fish.

(I) Normalized mean L-MO glial Ca^{2+} signals centered on the onset of passivity, induced by TRPV1 in closed loop (top) or by motor futility in open loop (bottom). Surges appear similar. Ca^{2+} in TRPV1+ animals did not saturate.

(J) Experimental design. Local optogenetic excitation of glial processes in fish expressing CoChR or Opto-α1-AR increases glial Ca^{2+} in L-MO (CoChR shown).

(K) Optogenetic (CoChR) excitation of L-MO glia reduced swimming for ~10 s. Blue light outside the brain caused a much milder and briefer reduction.

(L and M) Population data show reduced swimming relative to controls in fish expressing CoChR (L) or Opto-α1-AR (M) in radial astrocytes. Two-tailed paired t test, CoChR, p = 0.00012, n = 9 fish; Opto-α1-AR, p = 0.0065, n = 10 fish.

(N–P) Controls for (K)–(M) show no effect in fish without CoChR or Opto-α1-AR. Two-tailed paired t test, CoChR: p = 0.89, n = 6 fish. Opto-α1-AR, p = 0.50, n = 8 fish.

(Q) Summary of Figures 4 and 5, showing significant increases in passivity after increases in astrocytic Ca^{2+}, decreases after astrocytic silencing, and no effects in control experiments. Effects are normalized by maximal possible change (control passive % for decreases; control active % for increases) for each fish.

(R) Optogenetic activation of L-MO radial astrocytes activates GABAergic neurons at edge of L-MO in *Tg(gfap:CoChR-eGFP);Tg(elavl3:jRGECO1b)* fish. Left: maximum-intensity projection of $\Delta F/F$ after optogenetic activation of radial astrocytes (magenta) and anatomy (cyan) in example fish. Inset: Ca^{2+} in GABAergic neurons after astrocytic activation; stimulation time in cyan. Top right: *Tg(gad1b:RFP);Tg(elavl3:H2B-GCaMP6f)* used to label GABAergic neurons; bottom right: *Tg(gad1b:Gal4);Tg(UAS:CoChR-eGFP)* to sparsely label GABAergic neurons.

(S) Activation of GABAergic cells with laser light in *Tg(Gad1b:Gal4);Tg(UAS:CoChR-eGFP)* suppresses swimming. Two-tailed paired t test, p = 0.0074, n = 7 fish.

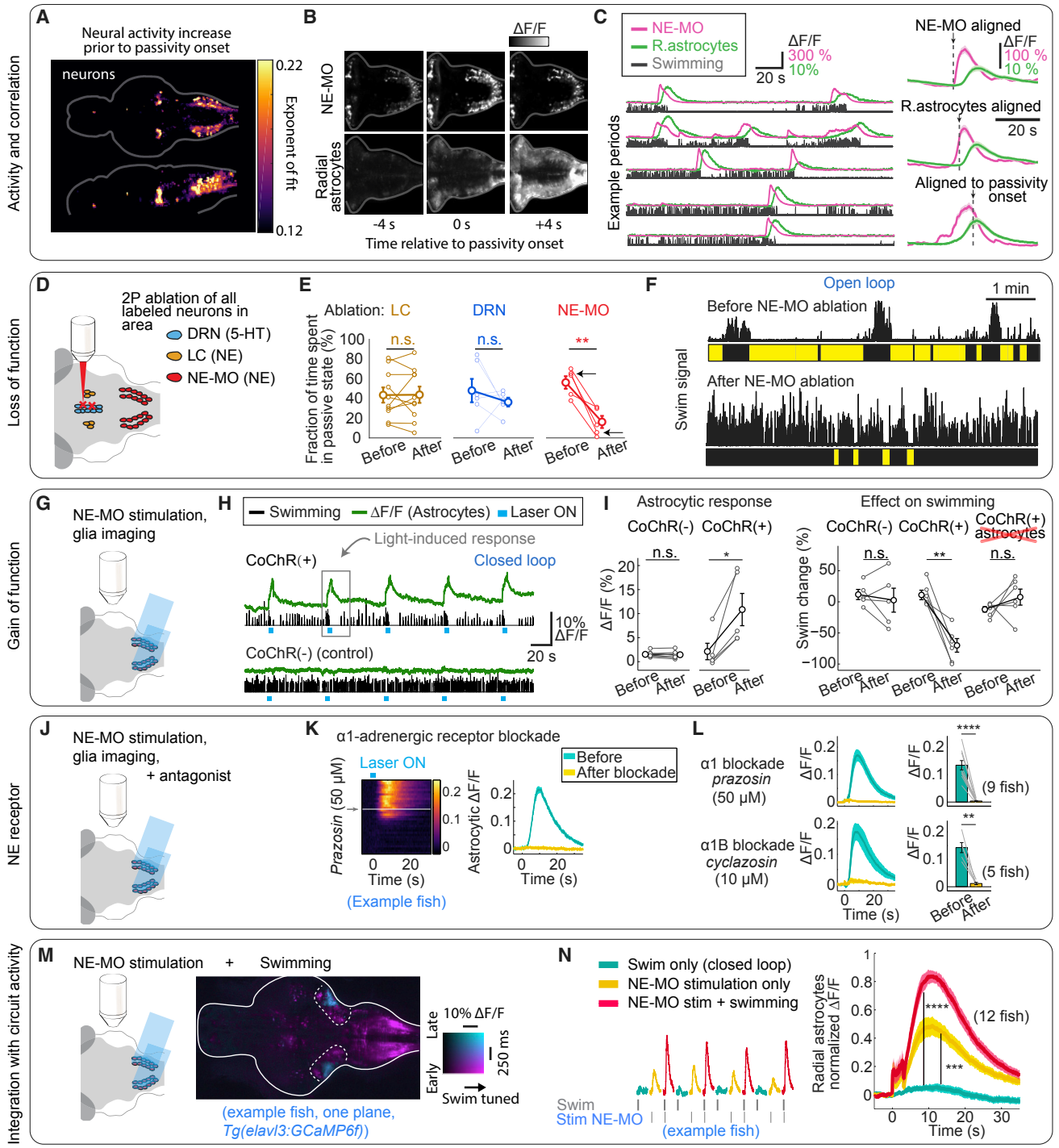


Figure 6. Noradrenergic System Activates Radial Astrocytes and Is Required for Futility-Induced Passivity

(A) Rate of neuronal activity increase before passivity onset quantified by the exponent of an exponential fit to average Ca^{2+} signals before passivity (overlay of 5 fish, *Tg(elavl3:GCaMP6f)*). Population with the highest rate of increase overlaps with NE neurons (see Figure S6A).

(B) Simultaneous imaging of NE neurons and radial astrocytes. Example Ca^{2+} signals just before, at, and after passivity onset.

(C) Examples of NE-MO activity, glial Ca^{2+} increases, and swimming in an alternating closed- and open-loop protocol. Right: average Ca^{2+} activity from NE-MO and radial astrocytes, aligned to NE-MO response onset, radial astrocyte response, or passivity onset.

(D) Experimental design: single-cell 2-photon ablation of LC, NE-MO, or DRN in transgenic zebrafish labeling these nuclei.

(E) Time in passivity was strongly reduced after NE-MO ablation, but not after ablation of LC or DRN. Two-tailed paired t test, p = 0.95, n = 10 fish for LC; p = 0.40, n = 6 fish for DRN; p = 0.0070, n = 5 fish for NE-MO.

(legend continued on next page)

What information does NE convey to radial astrocytes? If L-MO radial astrocytes encode swim futility (the time integral of failures), we hypothesized that the NE system encodes swim failures. To look for such signals, we imaged NE cells in three conditions, randomly interleaved over swim bouts: (1) fictive swimming in closed loop with backward visual flow to mimic "successful" swimming (92% of swims), (2) backward visual flow without swimming (4%), and (3) swimming in open loop without visual feedback (4%) to mimic "failed" swims (Figure 7A). If NE cells detect mismatch (Keller et al., 2012), they should be more activated in (3) than in (1) or (2). Indeed, that is what we found (Figure 7B): NE-MO responses were small or absent when fish swam with naturalistic feedback but large when they swam without feedback.

We next quantified whether NE-MO responses encode the difference between perceived and expected visual feedback during swimming. We randomized motosensory gain to elicit successful swims (gain > 0; swim generates backward visual motion) and failed swims (gain $\leq$ 0; swim with no visual change or motion opposite to expected; Figure 7C). NE-MO responses were consistent with a mismatch signal, increasing when motor vigor increased without visual effect (Figures 7D and 7E), or when vigor was constant and backward visual flow was absent or forward flow was seen (Figures 7F and 7G). Thus, NE-MO signals swim failures by encoding sensory-motor mismatch.

Radial Astrocytes Integrate Evidence that Actions Are Futile

Determining that a behavior is futile requires integrating the results of multiple successive attempts. Failed swims lead to elevated NE-MO activity, which in turn excites astrocytes (Figure 7H). NE-MO Ca^{2+} eventually leveled out or decreased (Figures 7H and 7I), while astrocytic Ca^{2+} continued to increase (Figures 7H and 7I). As radial astrocytes respond to sensory-motor mismatch signals from NE-MO and accumulate them over time, we asked whether they integrate evidence (Pinto et al., 2018; Shadlen and Newsome, 2001) that actions are futile. We designed a behavioral task in which swimming leads to visual flow probabilistically (Figure 7J). First, in closed loop, all swim bouts trigger visual flow. Next, every swim bout triggers visual flow with probability P. Glial Ca^{2+} rose fastest when P = 0 (always futile), did not rise when P = 1 (always effective), and increased at intermediate rates for intermediate P (Figure 7K). This shows that, following a switch from a closed-loop to a stochastic feedback protocol, radial astrocyte Ca^{2+} reflects evidence accumulation. Together, these results (Table S3) show mechanistically how radial astrocytes encode experience and control behavioral state.

DISCUSSION

Here, we found that the fish analog of the mammalian astrocyte is a central computational element of a circuit implementing a behavioral-state change after integrating sensory information. Specifically, radial astrocytes in a subregion of the brainstem in larval zebrafish temporally integrate noradrenergically encoded failures to accumulate evidence of futility before inducing a state of passivity ("giving up"). This behavioral pattern has a familiar combination of features: trying to achieve a goal, repeatedly failing despite trying harder, giving up temporarily, and then trying again—fish swim more vigorously in open loop (i.e., ineffective swim attempts), become passive, then swim again. Using whole-brain imaging, we found that astrocytic calcium was elevated just before—and during—passive states. Activation and silencing experiments established that these glial cells are required to trigger the passive state. Neuronal imaging and manipulation showed that the NE system encodes an expectation-outcome mismatch signal, which activates Ca^{2+} signaling in radial astrocytes. Thus, behavioral failures are detected by NE-MO and integrated by glia, which, after accumulating sufficient evidence of motor futility, trigger a passive behavioral state via GABAergic neurons. Once passivity is triggered, its persistence may be due to sustained effects of the glial cells on neurons, or a lag in the reactivation of swim circuits. These results imply that routinely including astrocytes in systems neuroscience studies may accelerate achieving a comprehensive understanding of brain function (Barres, 2008).

This study extends the known roles of glial cells to include a specific computation (temporal integration of a behaviorally relevant mismatch signal), resulting in a direct, fast effect on brain and behavioral state (induction of passivity within seconds). The traditional view of astrocytes as supporting cells for neurons is known to be incomplete (Araque et al., 2014; Kjaerby et al., 2017), as they actively influence neuronal activity.

(F) Example shows NE-MO ablation reduced futility-induced passivity. Arrows in (E), corresponding data points.

(G) Experimental design: optogenetic activation of NE-MO cells using DMD-targeted light delivery in fish expressing CoChR in NE cells while imaging radial astrocytes expressing jRGECO1b.

(H) Example fish, NE-MO activation led to increased L-MO glial Ca^{2+}, then passivity. Swimming increases briefly at light onset (see Figure S6D).

(I) Population data: NE-MO activation increased L-MO glial Ca^{2+} (left) and reduced swimming (right) in $CoChR^{+}$ fish. Two-tailed paired t test, $CoChR^{+}$ imaging, p = 0.025, n = 5 fish; $CoChR^{+}$ behavior, p = 0.0024, n = 6 fish; $CoChR^{-}$ imaging, p = 0.52, n = 7 fish; $CoChR^{-}$ behavior, p = 0.56, n = 5 fish; glial ablation, p = 0.27, n = 6 fish.

(J) Experiment to identify receptors. Experiment as in (G) with NE receptor antagonists added to bath.

(K) α1 NE receptor antagonist prazosin (50 μM) abolishes radial astrocyte response evoked by NE-MO stimulation.

(L) Population data. α1 NE receptor antagonist prazosin (50 μM) or α1B antagonist cyclazosin (10 μM) abolished radial astrocyte responses. Two-tailed paired t test, prazosin, $p = 4.6 \times 10^{-5}$, n = 9 fish; cyclazosin, p = 0.0042, n = 5 fish. Shading, SEM.

(M) Experimental design to test whether L-MO radial astrocytes integrate NE and local circuit activity. In closed loop, NE-MO cells are optogenetically stimulated during swimming (higher L-MO neuronal activity), or during not swimming.

(N) Integration of NE and local circuit activity. Radial astrocytic Ca^{2+} responses to NE during swimming are larger than to NE alone. Left: examples from successive trials. Right: summary of all fish. Two-tailed paired t test, $p = 2.4 \times 10^{-4}$ for NE-MO stimulation only versus swimming only, $p = 2.4 \times 10^{-8}$ for NE-MO stimulation + swimming versus swimming only, n = 12 fish. Shading, SEM.

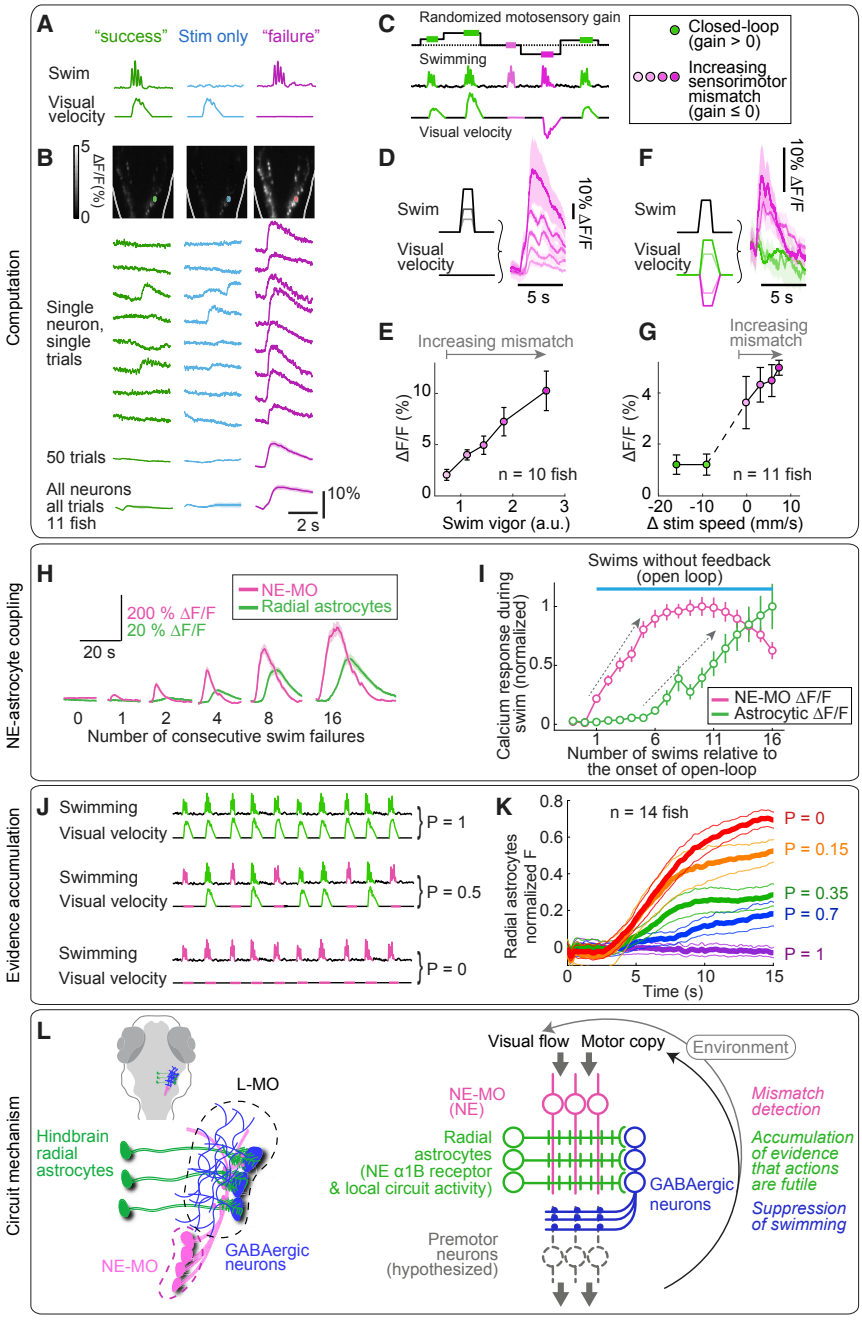


Figure 7. Noradrenergic System Signals Visuomotor Mismatch; Radial Astrocytes Accumulate Evidence that Actions Are Futile

For a Figure360 author presentation of this figure, see https://doi.org/10.1016/j.cell.2019.05.050. Figure360▷

(A) Mismatch assay. Left: in closed loop, motor output triggers visual feedback (motosensory gain > 0). Middle: stimulus only with no swimming. Right: sensory-motor mismatch in swim failures (gain = 0).

(B) Mismatch signals in NE-MO. In closed loop or visual signal alone, NE-MO responses were small. During swim failures, NE-MO responses were large. (See Figure S6H for LC, which does not respond to mismatch until next swim bout.)

(C) Extended mismatch assay: motosensory gain randomly set for each swim bout, varying combinations of motor vigor and visual velocity.

(D–G) NE-MO responses encode sensory-motor mismatch. (D and E) NE-MO responses grew with stronger motor vigor if no visual stimulus was shown. (D, example fish; E, population data, 10 fish.) (F and G) NE-MO responses were minimal in closed loop (motosensory gain > 0), whether gain was high or low. In open loop (no feedback, gain = 0) NE-MO responded. With opposite feedback (forward visual feedback during swimming, high mismatch, motosensory gain < 0), NE-MO response was larger. In (F and G), average motor vigor is similar across visual speeds due to randomization of gain. (F, example fish; G, population data, 11 fish). Shading, SEM.

(H) Radial astrocyte Ca^{2+} integrates swim failures. Fish swam in closed loop, then open loop for 1, 2, 4, 8, or 16 swim bouts, then back to closed loop. Glial signals increased with number of failed bouts after ~2 failed bouts; NE-MO signals began increasing with first failed bout. Trials with passivity not included. NE-MO signal sometimes saturated or decreased while glial Ca^{2+} continued to increase.

(I) Radial astrocyte Ca^{2+} accumulated with increasing numbers of swim failures. NE-MO Ca^{2+} also increased initially but saturated or decreased, while glial Ca^{2+} continued to increase with swim failures until feedback returns to closed loop. Gray arrows, accumulation.

(J) Experiment: information accumulation in radial astrocytes. In a swim bout, probability that fish sees visual feedback (backward visual flow) is P. P = 1, closed loop; P = 0, open loop; intermediate P gives probabilistic feedback.

(K) Information accumulation in radial astrocytes. Ca^{2+} signal, averaged over fish and trials, reflects P, the probability that a swim bout gives visual feedback, so glia accumulate information on swim futility over time (n = 14 fish).

(L) Circuit diagram. NE-MO represents mismatch signal, computed from visual and motor efference input. Noradrenergic axons from NE-MO excite glial processes in L-MO through α1B NE receptors (see Figure S6C for projections). Radial astrocytes integrate mismatch signals and suppress swimming via downstream GABAergic neurons.

In the hippocampus, astrocytes can be intermediates between cholinergic input and interneurons (Pabst et al., 2016). Glial cells in zebrafish are activated before and during generalized seizures (Verdugo et al., 2019). These and other studies (Adamsky et al., 2018; Ding et al., 2013; Paukert et al., 2014; Poskanzer and Yuste, 2016; Srinivasan et al., 2015) point to an integrated role for astrocytes in information processing in brain circuits.

Our study establishes radial astrocytes as essential computational elements in a circuit that mediates an adaptive behavioral response. The algorithm implemented by the radial astrocytes here is surprisingly analogous to that observed in mouse

hippocampal slices (Deemyad et al., 2018). There, astrocytes integrate neural signals over many seconds and then suddenly switch into a mode where they drive strong inhibition. This "integrate and suppress" functional motif may therefore generalize across neural systems and species. Interestingly, a role for astrocytes in regulating active and passive states of circuits, albeit in a mechanistically different form and on a longer timescale, was conjectured a century ago (Ramon y Cajal, 1895).

How do radial astrocytes drive activity in GABAergic edge cells? Astrocytes can influence synaptic transmission and neuronal firing in multiple ways (Bazargani and Attwell, 2016), but it is unclear which mechanisms are most relevant *in vivo* (Fiacco and McCarthy, 2018; Savtchouk and Volterra, 2018). In *Drosophila*, astrocytes inhibit downstream neurons likely through ATP secretion to influence chemotaxis and startle behavior (Ma et al., 2016). Astrocytes can release GABA (Angulo et al., 2008; Lee et al., 2010; Christensen et al., 2018). They can also influence neuronal firing by modulating extracellular ionic composition. On slow timescales, astrocytes affect neuronal firing via changes in the density of inward-rectifying K^+ membrane channels during depression in rodents (Cui et al., 2018). On faster timescales, glia-like support cells in the developing cochlea trigger neural activity in adjacent hair cells by modulating extracellular K^+ through a calcium-activated chloride channel (Wang et al., 2015). Our study shows that radial astrocytes modulate neurons and influence behavior over seconds (likely triggered by intracellular Ca^{2+} signaling, but possibly via other second messenger systems). The molecular mechanisms were not identified, but may include active control of extracellular ion concentrations near the GABAergic neurons. Radial astrocytes may affect behavior through more pathways than only the GABAergic L-MO cells. Other routes may exist between the glial cells and other neuron types, potentially including direct suppressive effects (Christensen et al., 2018; Wang et al., 2012) on premotor neurons.

Homology between radial astrocytes in zebrafish and astrocytes in mammals is suggested by several similarities. Both cell types have a bushy morphology, which in zebrafish appears after intense neurogenesis (2–3 days post fertilization [David Lyons, personal communication]), qualitatively similar to the emergence of bushy processes when mammalian radial glia convert to astrocytes at the end of embryonic development (Kriegstein and Alvarez-Buylla, 2009). We also found that fish radial astrocytes integrate norepinephrine with local circuit activity, similar to astrocytes in mammalian cortex (Paukert et al., 2014). These findings, in combination with overlapping gene expression and morphological features (Lyons and Talbot, 2014), suggest that the mechanisms shown here may be relevant for neuron-astrocyte interactions and glial control of behavior in other vertebrates.

Neuronal manipulation experiments come with caveats (Otchy et al., 2015); glial manipulation experiments come with more (see Table S1), in part because tools for manipulating glia are less established. Thus, we did six experiments to activate or silence radial astrocytes, attempting to falsify our emerging model that astrocytic activation halts swimming in futility-induced passivity. All six support a causal role for glia in triggering passivity.

Our finding that the NE system encodes the magnitude of failure is consistent with noradrenergic neurons encoding changes in relationships between actions and outcomes (Aston-Jones and Cohen, 2005; Bouret and Sara, 2004; Sara, 2009; Tervo et al., 2014). NE also encodes alertness in zebrafish (Lovett-Barron et al., 2017), which may correlate to the temporary increase in swim vigor after actions become ineffective.

Different action-outcome contingencies lead to different behaviors. When desired outcomes are attainable, motor learning fine-tunes actions; when they are unattainable, futility-induced passivity interrupts behavior. The serotonergic system mediates a form of motor learning in zebrafish by integrating backward visual flow during swimming (Kawashima et al., 2016). Thus, multiple neuromodulatory systems underlie complementary adaptive strategies: the serotonergic system encodes the degree of success of actions and leads to fine-tuning of swim vigor, while the noradrenergic system encodes failure and can drive passivity.

Futility-induced passivity belongs to a class of behaviors in which motor output is reduced, including learned helplessness (Maier, 1984), passive coping (Andalman et al., 2019; Warden et al., 2012), feigned death (Humphreys and Ruxton, 2018), and break-point behavior, where attempts to achieve reward are abandoned (Salamone et al., 2016). Although the behavior studied here is mediated by cells in the brainstem, this broader class of behaviors can involve other brain regions in mammals, including multiple neuromodulatory systems and cortical areas (Roseberry and Kreitzer, 2017).

Evidence-accumulation signals have been found in neuronal firing rates in multiple decision-making paradigms (Scott et al., 2017; Shadlen and Newsome, 2001). Here, we found that, following a switch from a closed-loop to a stochastic feedback protocol, calcium levels in radial astrocytes reflect evidence that actions are futile. The intracellular messengers in the pathway from NE receptors to intracellular Ca^{2+} increases may also be involved in the accumulation process. It remains to be elucidated which of these molecules gets thresholded (and how) to mediate the "decision" to trigger passivity.

We ascribe evidence accumulation to radial astrocytes and sensory-motor-mismatch encoding to NE-MO, but the boundary between the two operations may not be strict. NE-MO may take part in the accumulation, for example over shorter timescales or during the first few swim failures.

What class of computations might astrocytes support? In addition to having different inputs from neurons, these cell types may be specialized for different input-output transformations, based on their spatiotemporal dynamics. Neurons convert chemical inputs into membrane voltage, electrically integrate across thousands of such inputs, and fire action potentials in milliseconds. Although glial processes can respond to neurons in ~100 ms (Stobart et al., 2018), glia often operate on slower timescales using molecular signaling. Such properties may favor astrocytes for computations involving slow integration of activity. The molecular mechanisms of astrocytic computation are not yet clear, and understanding their spatial interactions with neurons may require incorporating astrocytes into electron microscopy-based circuit reconstruction (Briggman and Bock, 2012). Broadening such efforts to include astrocytes will likely accelerate neuroscience discovery.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Zebrafish
  - Transgenesis
- METHOD DETAILS
  - Preparation for fictive-behavior assays
  - Fictive behavior in virtual reality
  - Futility-induced passivity behavioral analysis
  - DMD-based optogenetic stimulation
  - Two-photon ablation of cells
  - Forebrain lesioning
  - TRPV1-mediated activation of radial astrocytes
  - IP_3 receptor blockade
  - Analysis of fluorescence imaging data
  - Confocal microscopy
  - Single-cell radial astrocyte tracing
  - *In situ* hybridization
  - Immunohistochemistry with expansion microscopy
  - RNA sequencing
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2019.05.050.

A video abstract is available at https://doi.org/10.1016/j.cell.2019.05.050#mmc11.

ACKNOWLEDGMENTS

We thank C. Wang, P. Keller, and T. Kawashima for help with optics; A. Lemire and L. Wang for help with transcriptomics; P. Tillberg and A. Hu for help with anatomy imaging and tissue expansion; M. Copeland for help with *in situ* hybridization; M. Koyama for help with electrophysiology infrastructure; I. Bianco, H. Burgess, and D. Lyons for suggestions; M. Koyama, N. Spruston, D. Bergles, J. Du, J. Lippincott-Schwartz, J. Freeman, J. Wittenbach, D. Wolpert, J. Dudman, M. Hale, E. Yaksi, C. Brody, F. Engert, K. Monk, M. Hatten, B. Khakh, Y. Liao, and C. Zuker for discussions; V. Jayaraman, N. Li, A. Lee, U. Heberlein, W. Ge, Y. Wang, J. Phillips, S. Chen, D. Stern, D. Lyons, D. Schoppik, S. Liddelow, M. Warden, and S. Chen for comments on the paper; J. Rouchard, J. Cox and vivarium staff for fish husbandry and help making transgenic fish; D. Prober for the TRPV1 gene; P. Raymond for the *gfap* promoter; and E. Boyden for the CoChR gene. This work was supported by the Howard Hughes Medical Institute and Simons Collaboration on the Global Brain Awards (325171, 542943SPI).

AUTHOR CONTRIBUTIONS

Y.M., D.V.B., M.R., S.N., and M.B.A. conceived the project. Y.M., D.V.B., and S.N. performed the experiments. M.R. created the analysis pipeline. M.R., Y.M., and D.V.B. analyzed the data. D.V.B. built the optical excitation setup. S.N. and C.-T.Y. generated transgenic fish. M.T. and Y.M. did glial electrophysiology. L.L.L. analyzed RNA-seq data. M.B.A. supervised the study. Y.M., D.B., M.R., B.D.M., and M.B.A. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 3, 2018
Revised: January 31, 2019
Accepted: May 23, 2019
Published: June 20, 2019

REFERENCES

Adamsky, A., Kol, A., Kreisel, T., Doron, A., Ozeri-Engelhard, N., Melcer, T., Refaeli, R., Horn, H., Regev, L., Groysman, M., et al. (2018). Astrocytic Activation Generates De Novo Neuronal Potentiation and Memory Enhancement. Cell *174*, 59–71.e14.

Ahrens, M.B., Li, J.M., Orger, M.B., Robson, D.N., Schier, A.F., Engert, F., and Portugues, R. (2012). Brain-wide neuronal dynamics during motor adaptation in zebrafish. Nature *485*, 471–477.

Ahrens, M.B., Orger, M.B., Robson, D.N., Li, J.M., and Keller, P.J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. Nat. Methods *10*, 413–420.

Airan, R.D., Thompson, K.R., Fenno, L.E., Bernstein, H., and Deisseroth, K. (2009). Temporally precise in vivo control of intracellular signalling. Nature *458*, 1025–1029.

Andalman, A.S., Burns, V.M., Lovett-Barron, M., Broxton, M., Poole, B., Yang, S.J., Grosenick, L., Lerner, T.N., Chen, R., Benster, T., et al. (2019). Neuronal Dynamics Regulating Brain and Behavioral State Transitions. Cell *177*, 970–985.e20.

Angulo, M.C., Le Meur, K., Kozlov, A.S., Charpak, S., and Audinat, E. (2008). GABA, a forgotten gliotransmitter. Prog. Neurobiol. *86*, 297–303.

Anthony, T.E., Klein, C., Fishell, G., and Heintz, N. (2004). Radial glia serve as neuronal progenitors in all regions of the central nervous system. Neuron *41*, 881–890.

Araque, A., Carmignoto, G., Haydon, P.G., Oliet, S.H.R., Robitaille, R., and Volterra, A. (2014). Gliotransmitters travel in time and space. Neuron *81*, 728–739.

Aston-Jones, G., and Cohen, J.D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu Rev Neurosci. *28*, 403–450.

Avants, B.B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., and Gee, J.C. (2010). The optimal template effect in hippocampus studies of diseased populations. Neuroimage *49*, 2457–2466.

Banker, G.A. (1980). Trophic interactions between astroglial cells and hippocampal neurons in culture. Science *209*, 809–810.

Barres, B.A. (2008). The mystery and magic of glia: a perspective on their roles in health and disease. Neuron *60*, 430–440.

Bazargani, N., and Attwell, D. (2016). Astrocyte calcium signaling: the third wave. Nat. Neurosci. *19*, 182–189.

Bekar, L.K., He, W., and Nedergaard, M. (2008). Locus coeruleus α-adrenergic-mediated activation of cortical astrocytes in vivo. Cereb. Cortex *18*, 2789–2795.

Bernardos, R.L., and Raymond, P.A. (2006). GFAP transgenic zebrafish. Gene Expr. Patterns *6*, 1007–1013.

Berry, M.W., Browne, M., Langville, A.N., Pauca, V.P., and Plemmons, R.J. (2007). Algorithms and applications for approximate nonnegative matrix factorization. Comput. Stat. Data Anal. *52*, 155–173.

Bouret, S., and Sara, S.J. (2004). Reward expectation, orientation of attention and locus coeruleus-medial frontal cortex interplay during learning. Eur. J. Neurosci. *20*, 791–802.

Boutsidis, C., and Gallopoulos, E. (2008). SVD based initialization: A head start for nonnegative matrix factorization. Pattern Recognit. *41*, 1350–1362.

Briggman, K.L., and Bock, D.D. (2012). Volume electron microscopy for neuronal circuit reconstruction. Curr. Opin. Neurobiol. *22*, 154–161.

Cembrowski, M.S., Wang, L., Sugino, K., Shields, B.C., and Spruston, N. (2016). Hipposeq: a comprehensive RNA-seq database of gene expression in hippocampal principal neurons. eLife *5*, e14997.

Chen, T.-W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295–300.

Chen, S., Chiu, C.N., McArthur, K.L., Fetcho, J.R., and Prober, D.A. (2016). TRP channel mediated neuronal activation and ablation in freely behaving zebrafish. Nat. Methods *13*, 147–150.

Christensen, R.K., Delgado-Lezama, R., Russo, R.E., Lind, B.L., Alcocer, E.L., Rath, M.F., Fabbiani, G., Schmitt, N., Lauritzen, M., Petersen, A.V., et al. (2018). Spinal dorsal horn astrocytes release GABA in response to synaptic activation. J. Physiol. *596*, 4983–4994.

Cui, Y., Yang, Y., Ni, Z., Dong, Y., Cai, G., Foncelle, A., Ma, S., Sang, K., Tang, S., Li, Y., et al. (2018). Astroglial Kir4.1 in the lateral habenula drives neuronal bursts in depression. Nature *554*, 323–327.

Cuoghi, B., and Mola, L. (2009). Macroglial cells of the teleost central nervous system: a survey of the main types. Cell Tissue Res. *338*, 319–332.

Dana, H., Mohar, B., Sun, Y., Narayan, S., Gordus, A., Hasseman, J.P., Tsegaye, G., Holt, G.T., Hu, A., Walpita, D., et al. (2016). Sensitive red protein calcium indicators for imaging neural activity. eLife *5*, e12727.

Deemyad, T., Lüthi, J., and Spruston, N. (2018). Astrocytes integrate and drive action potential firing in inhibitory subnetworks. Nat. Commun. *9*, 4336.

Ding, F., O'Donnell, J., Thrane, A.S., Zeppenfeld, D., Kang, H., Xie, L., Wang, F., and Nedergaard, M. (2013). α1-Adrenergic receptors mediate coordinated Ca2+ signaling of cortical astrocytes in awake, behaving mice. Cell Calcium *54*, 387–394.

Dunn, T.W., Mu, Y., Narayan, S., Randlett, O., Naumann, E.A., Yang, C.-T., Schier, A.F., Freeman, J., Engert, F., and Ahrens, M.B. (2016). Brain-wide mapping of neural activity controlling zebrafish exploratory locomotion. eLife *5*, e12741.

Farrar, M.J., Kolkman, K.E., and Fetcho, J.R. (2018). Features of the structure, development, and activity of the zebrafish noradrenergic system explored in new CRISPR transgenic lines. J. Comp. Neurol. *526*, 2493–2508.

Fiacco, T.A., and McCarthy, K.D. (2018). Multiple Lines of Evidence Indicate That Gliotransmission Does Not Occur under Physiological Conditions. J. Neurosci. *38*, 3–13.

Figueiredo, M., Lane, S., Stout, R.F., Jr., Liu, B., Parpura, V., Teschemacher, A.G., and Kasparov, S. (2014). Comparative analysis of optogenetic actuators in cultured astrocytes. Cell Calcium *56*, 208–214.

Freeman, J., Vladimirov, N., Kawashima, T., Mu, Y., Sofroniew, N.J., Bennett, D.V., Rosen, J., Yang, C.-T., Looger, L.L., and Ahrens, M.B. (2014). Mapping brain activity at scale with cluster computing. Nat. Methods *11*, 941–950.

Freifeld, L., Odstrcil, I., Förster, D., Ramirez, A., Gagnon, J.A., Randlett, O., Costa, E.K., Asano, S., Celiker, O.T., Gao, R., et al. (2017). Expansion microscopy of zebrafish for neuroscience and developmental biology studies. Proc. Natl. Acad. Sci. USA *114*, E10799–E10808.

Fujii, Y., Maekawa, S., and Morita, M. (2017). Astrocyte calcium waves propagate proximally by gap junction and distally by extracellular diffusion of ATP released from volume-regulated anion channels. Sci. Rep. *7*, 13115.

Fujita, T., Chen, M.J., Li, B., Smith, N.A., Peng, W., Sun, W., Toner, M.J., Kress, B.T., Wang, L., Benraiss, A., et al. (2014). Neuronal transgene expression in dominant-negative SNARE mice. J. Neurosci. *34*, 16594–16604.

Gafni, J., Munsch, J.A., Lam, T.H., Catlin, M.C., Costa, L.G., Molinski, T.F., and Pessah, I.N. (1997). Xestospongins: potent membrane permeable blockers of the inositol 1,4,5-trisphosphate receptor. Neuron *19*, 723–733.

Gourine, A.V., Kasymov, V., Marina, N., Tang, F., Figueiredo, M.F., Lane, S., Teschemacher, A.G., Spyer, K.M., Deisseroth, K., and Kasparov, S. (2010). Astrocytes control breathing through pH-dependent release of ATP. Science *329*, 571–575.

Gross-Thebing, T., Paksa, A., and Raz, E. (2014). Simultaneous high-resolution detection of multiple transcripts combined with localization of proteins in whole-mount embryos. BMC Biol. *12*, 55.

Grupp, L., Wolburg, H., and Mack, A.F. (2010). Astroglial structures in the zebrafish brain. J. Comp. Neurol. *518*, 4277–4287.

Haydon, P.G., and Nedergaard, M. (2014). How do astrocytes participate in neural plasticity? Cold Spring Harb. Perspect. Biol. *7*, a020438.

Hempel, C.M., Sugino, K., and Nelson, S.B. (2007). A manual method for the purification of fluorescently labeled neurons from the mammalian brain. Nat. Protoc. *2*, 2924–2929.

Hoyer, P.O. (2004). Non-negative matrix factorization with sparseness constraints. J. Mach. Learn. Res. *5*, 1457–1469.

Humphreys, R.K., and Ruxton, G.D. (2018). A review of thanatosis (death feigning) as an anti-predator behaviour. Behav. Ecol. Sociobiol. (Print) *72*, 22.

Kawashima, T., Zwart, M.F., Yang, C.-T., Mensh, B.D., and Ahrens, M.B. (2016). The Serotonergic System Tracks the Outcomes of Actions to Mediate Short-Term Motor Learning. Cell *167*, 933–946.e20.

Keller, G.B., Bonhoeffer, T., and Hübener, M. (2012). Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. Neuron *74*, 809–815.

Kimura, Y., Hisano, Y., Kawahara, A., and Higashijima, S. (2014). Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. Sci. Rep. *4*, 6545.

Kjaerby, C., Rasmussen, R., Andersen, M., and Nedergaard, M. (2017). Does Global Astrocytic Calcium Signaling Participate in Awake Brain State Transitions and Neuronal Circuit Function? Neurochem. Res. *42*, 1810–1822.

Klapoetke, N.C., Murata, Y., Kim, S.S., Pulver, S.R., Birdsey-Benson, A., Cho, Y.K., Morimoto, T.K., Chuong, A.S., Carpenter, E.J., Tian, Z., et al. (2014). Independent optical excitation of distinct neural populations. Nat. Methods *11*, 338–346.

Koyama, M., Kinkhabwala, A., Satou, C., Higashijima, S., and Fetcho, J. (2011). Mapping a sensory-motor network onto a structural and functional ground plan in the hindbrain. Proc Natl Acad Sci USA *108*, 1170–1175.

Kriegstein, A., and Alvarez-Buylla, A. (2009). The glial nature of embryonic and adult neural stem cells. Annu. Rev. Neurosci. *32*, 149–184.

Lee, S., Yoon, B.-E., Berglund, K., Oh, S.-J., Park, H., Shin, H.-S., Augustine, G.J., and Lee, C.J. (2010). Channel-mediated tonic GABA release from glia. Science *330*, 790–796.

Li, J., Zhang, B.B., Ren, Y.G., Gu, S.Y., Xiang, Y.H., and Du, J.L. (2015). Intron targeting-mediated and endogenous gene integrity-maintaining knockin in zebrafish using the CRISPR/Cas9 system. Cell Res. *25*, 634–637.

Liao, J.C., and Fetcho, J.R. (2008). Shared versus specialized glycinergic spinal interneurons in axial motor circuits of larval zebrafish. J. Neurosci. *28*, 12982–12992.

Lovett-Barron, M., Andalman, A.S., Allen, W.E., Vesuna, S., Kauvar, I., Burns, V.M., and Deisseroth, K. (2017). Ancestral Circuits for the Coordinated Modulation of Brain State. Cell *171*, 1411–1423.e17.

Lyons, D.A., and Talbot, W.S. (2014). Glial cell development and function in zebrafish. Cold Spring Harb. Perspect. Biol. *7*, a020586.

Ma, Z., Stork, T., Bergles, D.E., and Freeman, M.R. (2016). Neuromodulators signal through astrocytes to alter neural circuit activity and behaviour. Nature *539*, 428–432.

Maier, S.F. (1984). Learned helplessness and animal models of depression. Prog. Neuropsychopharmacol. Biol. Psychiatry *8*, 435–446.

Marquart, G.D., Tabor, K.M., Brown, M., Strykowski, J.L., Varshney, G.K., LaFave, M.C., Mueller, T., Burgess, S.M., Higashijima, S., and Burgess, H.A. (2015). A 3D Searchable Database of Transgenic Zebrafish Gal4 and Cre Lines for Functional Neuroanatomy Studies. Front. Neural Circuits *9*, 78.

Morquette, P., Verdier, D., Kadala, A., Féthière, J., Philippe, A.G., Robitaille, R., and Kolta, A. (2015). An astrocyte-dependent mechanism for neuronal rhythmogenesis. Nat. Neurosci. *18*, 844–854.

Munns, C.H., Chung, M.-K., Sanchez, Y.E., Amzel, L.M., and Caterina, M.J. (2015). Role of the outer pore domain in transient receptor potential vanilloid 1 dynamic permeability to large cations. J. Biol. Chem. *290*, 5707–5724.

Nedergaard, M. (1994). Direct signaling from astrocytes to neurons in cultures of mammalian brain cells. Science *263*, 1768–1771.

Orger, M.B., Smear, M.C., Anstis, S.M., and Baier, H. (2000). Perception of Fourier and non-Fourier motion by larval zebrafish. Nat. Neurosci. *3*, 1128–1133.

Otchy, T.M., Wolff, S.B.E., Rhee, J.Y., Pehlevan, C., Kawai, R., Kempf, A., Gobes, S.M.H., and Ölveczky, B.P. (2015). Acute off-target effects of neural circuit manipulations. Nature *528*, 358–363.

Pabst, M., Braganza, O., Dannenberg, H., Hu, W., Pothmann, L., Rosen, J., Mody, I., van Loo, K., Deisseroth, K., Becker, A.J., et al. (2016). Astrocyte Intermediaries of Septal Cholinergic Modulation in the Hippocampus. Neuron *90*, 853–865.

Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. Nature *435*, 814–818.

Panier, T., Romano, S.A., Olive, R., Pietri, T., Sumbre, G., Candelier, R., and Debrégeas, G. (2013). Fast functional imaging of multiple brain regions in intact zebrafish larvae using selective plane illumination microscopy. Front. Neural Circuits *7*, 65.

Park, H.C., Kim, C.H., Bae, Y.K., Yeo, S.Y., Kim, S.H., Hong, S.K., Shin, J., Yoo, K.W., Hibi, M., Hirano, T., et al. (2000). Analysis of upstream elements in the HuC promoter leads to the establishment of transgenic zebrafish with fluorescent neurons. Dev. Biol. *227*, 279–293.

Parpura, V., Basarsky, T.A., Liu, F., Jeftinija, K., Jeftinija, S., and Haydon, P.G. (1994). Glutamate-mediated astrocyte-neuron signalling. Nature *369*, 744–747.

Paukert, M., Agarwal, A., Cha, J., Doze, V.A., Kang, J.U., and Bergles, D.E. (2014). Norepinephrine controls astroglial responsiveness to local circuit activity. Neuron *82*, 1263–1270.

Pinto, L., Koay, S.A., Engelhard, B., Yoon, A.M., Deverett, B., Thiberge, S.Y., Witten, I.B., Tank, D.W., and Brody, C.D. (2018). An Accumulation-of-Evidence Task Using Visual Pulses for Mice Navigating in Virtual Reality. Front. Behav. Neurosci. *12*, 36.

Portugues, R., and Engert, F. (2011). Adaptive locomotor behavior in larval zebrafish. Front. Syst. Neurosci. *5*, 72.

Poskanzer, K.E., and Yuste, R. (2016). Astrocytes regulate cortical state switching in vivo. Proc. Natl. Acad. Sci. USA *113*, E2675–E2684.

Ramon y Cajal, S. (1895). Algunas conjeturas sobre el mecanismo anatómico de la ideación, asociación y atención. Rev. Med. Cir. Pract. *36*, 497–508.

Randlett, O., Wee, C.L., Naumann, E.A., Nnaemeka, O., Schoppik, D., Fitzgerald, J.E., Portugues, R., Lacoste, A.M.B., Riegler, C., Engert, F., and Schier, A.F. (2015). Whole-brain activity mapping onto a zebrafish brain atlas. Nat. Methods *12*, 1039–1046.

Rinaman, L. (2011). Hindbrain noradrenergic A2 neurons: diverse roles in autonomic, endocrine, cognitive, and behavioral functions. Am. J. Physiol. Regul. Integr. Comp. Physiol. *300*, R222–R235.

Rock, I., and Smith, D. (1986). The optomotor response and induced motion of the self. Perception *15*, 497–502.

Roseberry, T., and Kreitzer, A. (2017). Neural circuitry for behavioural arrest. Philos. Trans. R. Soc. B Biol. Sci. *372*, 20160197.

Salamone, J.D., Correa, M., Yohn, S., Lopez Cruz, L., San Miguel, N., and Alatorre, L. (2016). The pharmacology of effort-related choice behavior: Dopamine, depression, and individual differences. Behav. Processes *127*, 3–17.

Sanz-Salvador, L., Andrés-Borderia, A., Ferrer-Montiel, A., and Planells-Cases, R. (2012). Agonist- and Ca2+-dependent desensitization of TRPV1 channel targets the receptor to lysosomes for degradation. J. Biol. Chem. *287*, 19462–19471.

Sara, S.J. (2009). The locus coeruleus and noradrenergic modulation of cognition. Nat. Rev. Neurosci. *10*, 211–223.

Satou, C., Kimura, Y., Hirata, H., Suster, M.L., Kawakami, K., and Higashijima, S. (2013). Transgenic tools to characterize neuronal properties of discrete populations of zebrafish neurons. Development *140*, 3927–3931.

Savtchouk, I., and Volterra, A. (2018). Gliotransmission: Beyond Black-and-White. J. Neurosci. *38*, 14–25.

Schummers, J., Yu, H., and Sur, M. (2008). Tuned responses of astrocytes and their influence on hemodynamic signals in the visual cortex. Science *320*, 1638–1643.

Scott, B.B., Constantinople, C.M., Akrami, A., Hanks, T.D., Brody, C.D., and Tank, D.W. (2017). Fronto-parietal Cortical Circuits Encode Accumulated Evidence with a Diversity of Timescales. Neuron *95*, 385–398.e5.

Shadlen, M.N., and Newsome, W.T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J. Neurophysiol. *86*, 1916–1936.

Slezak, M., Kandler, S., Van Veldhoven, P.P., Bonin, V., and Holt, M.G. (2018). Astrocytes integrate local sensory and brain-wide neuromodulatory signals. bioRxiv. https://doi.org/10.1101/381434.

Sloan, S.A., and Barres, B.A. (2014). Looks can be deceiving: reconsidering the evidence for gliotransmission. Neuron *84*, 1112–1115.

Srinivasan, R., Huang, B.S., Venugopal, S., Johnston, A.D., Chai, H., Zeng, H., Golshani, P., and Khakh, B.S. (2015). Ca(2+) signaling in astrocytes from Ip3r2(-/-) mice in brain slices and during startle responses in vivo. Nat. Neurosci. *18*, 708–717.

Stobart, J.L., Ferrari, K.D., Barrett, M.J.P., Glück, C., Stobart, M.J., Zuend, M., and Weber, B. (2018). Cortical Circuit Activity Evokes Rapid Astrocyte Calcium Signals on a Similar Timescale to Neurons. Neuron *98*, 726–735.e4.

Tabor, K.M., Bergeron, S.A., Horstick, E.J., Jordan, D.C., Aho, V., Porkka-Heiskanen, T., Haspel, G., and Burgess, H.A. (2014). Direct activation of the Mauthner cell by electric field pulses drives ultrarapid escape responses. J. Neurophysiol. *112*, 834–844.

Tay, T.L., Ronneberger, O., Ryu, S., Nitschke, R., and Driever, W. (2011). Comprehensive catecholaminergic projectome analysis reveals single-neuron integration of zebrafish ascending and descending dopaminergic systems. Nat. Commun. *2*, 171.

Tervo, D.G.R., Proskurin, M., Manakov, M., Kabra, M., Vollmer, A., Branson, K., and Karpova, A.Y. (2014). Behavioral variability through stochastic choice and its gating by anterior cingulate cortex. Cell *159*, 21–32.

Tillberg, P.W., Chen, F., Piatkevich, K.D., Zhao, Y., Yu, C.-C.J., English, B.P., Gao, L., Martorell, A., Suk, H.-J., Yoshida, F., et al. (2016). Protein-retention expansion microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies. Nat. Biotechnol. *34*, 987–992.

Uchida, D., Yamashita, M., Kitano, T., and Iguchi, T. (2002). Oocyte apoptosis during the transition from ovary-like tissue to testes during sex differentiation of juvenile zebrafish. J. Exp. Biol. *205*, 711–718.

Urasaki, A. (2009). Analysis of genes and genome by the tol2-mediated gene and enhancer trap methods. Methods Mol Biol. *546*, 85–102.

Verdugo, C.D., Myren-Svelstad, S., Deneubourg, C., Pelgrims, R., Muto, A., Kawakami, K., Jurisch-Yaksi, N., and Yaksi, E. (2019). Glia-neuron interactions underlie state transitions to generalized seizures. bioRxiv. https://doi.org/10.1101/509521.

Verkhratsky, A., Orkand, R.K., and Kettenmann, H. (1998). Glial calcium: homeostasis and signaling function. Physiol. Rev. *78*, 99–141.

Vladimirov, N., Mu, Y., Kawashima, T., Bennett, D.V., Yang, C.-T., Looger, L.L., Keller, P.J., Freeman, J., and Ahrens, M.B. (2014). Light-sheet functional imaging in fictively behaving zebrafish. Nat. Methods *11*, 883–884.

Vladimirov, N., Wang, C., Höckendorf, B., Pujala, A., Tanimoto, M., Mu, Y., Yang, C.-T., Wittenbach, J.D., Freeman, J., Preibisch, S., et al. (2018). Brain-wide circuit interrogation at the cellular level guided by online analysis of neuronal function. Nat. Methods *15*, 1117–1125.

Wang, F., Smith, N.A., Xu, Q., Fujita, T., Baba, A., Matsuda, T., Takano, T., Bekar, L., and Nedergaard, M. (2012). Astrocytes modulate neural network activity by Ca^2+-dependent uptake of extracellular K+. Sci. Signal. *5*, ra26.

Wang, H.C., Lin, C.-C., Cheung, R., Zhang-Hooks, Y., Agarwal, A., Ellis-Davies, G., Rock, J., and Bergles, D.E. (2015). Spontaneous Activity of Cochlear Hair Cells Triggered by Fluid Secretion Mechanism in Adjacent Support Cells. Cell *163*, 1348–1359.

Warden, M.R., Selimbeyoglu, A., Mirzabekov, J.J., Lo, M., Thompson, K.R., Kim, S.-Y., Adhikari, A., Tye, K.M., Frank, L.M., and Deisseroth, K. (2012).

A prefrontal cortex-brainstem neuronal projection that controls response to behavioural challenge. Nature *492*, 428–432.

Westerfield, M. (2007). A Guide for the Laboratory Use of Zebrafish (Danio Rerio).

White, R.M., Sessa, A., Burke, C., Bowman, T., LeBlanc, J., Ceol, C., Bourque, C., Dovey, M., Goessling, W., Burns, C.E., and Zon, L.I. (2008). Transparent adult zebrafish as a tool for in vivo transplantation analysis. Cell Stem Cell *2*, 183–189.

Yu, X., Taylor, A.M.W., Nagai, J., Golshani, P., Evans, C.J., Coppola, G., and Khakh, B.S. (2018). Reducing Astrocyte Calcium Signaling In Vivo Alters Striatal Microcircuits and Causes Repetitive Behavior. Neuron *99*, 1170–1187.e9.

Zaharia, M., Chowdhury, M., Franklin, M.J., Shenker, S., and Stoica, I. (2010). Spark: Cluster computing with working sets. Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing *10*.

Zhu, P., Fajardo, O., Shum, J., Zhang Schärer, Y.-P., and Friedrich, R.W. (2012a). High-resolution optical control of spatiotemporal neuronal activity patterns in zebrafish using a digital micromirror device. Nat. Protoc. *7*, 1410–1425.

Zhu, S., Lee, J.-S., Guo, F., Shin, J., Perez-Atayde, A.R., Kutok, J.L., Rodig, S.J., Neuberg, D.S., Helman, D., Feng, H., et al. (2012b). Activated ALK collaborates with MYCN in neuroblastoma pathogenesis. Cancer Cell *21*, 362–373.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-α2B- adrenergic receptor	Life Technologies	Cat# MA1-12721; RRID:AB_2225066
Bacterial and Virus Strains		
dbh-GFP-pDEST-ISce	(Zhu et al., 2012b)	N/A
FCK-gene86-GFP (CoChR)	(Klapoetke et al., 2014)	N/A
(Tol2-Isl-gal4vp16-4; UAS:TrpV1-tagRFpt)	(Chen et al., 2016)	N/A
pEGFP-gfap (Intron1/5′/Exon1-zebrafish)	Addgene	Cat# 39761
pcDNA3.1-opto-α1AR-EYFP	Addgene	Cat# 20947
Experimental Models: Organisms/Strains		
Tg(elavl3:GCaMP6f)^{jf1}	Janelia Research Campus (JRC) (Freeman et al., 2014)	N/A
Tg(elavl3:jRGECO1b)^{jf17}	JRC (Dana et al., 2016)	N/A
Tg(elavl3:H2B-GCaMP6f)^{jf7}	JRC (Dunn et al., 2016)	N/A
Tg(gfap:jRGECO1b)^{jf33}	JRC	N/A
Tg(gfap:GCaMP6f)^{jf22}	JRC	N/A
Tg(gfap:CoChR-eGFP)^{jf29}	JRC	N/A
Tg(gfap:opto-α1AR-eYFP)^{jf81}	JRC	N/A
Tg(gfap:H2B-GCaMP6f)^{jf23}	JRC	N/A
Tg(dbh:KalTA4)^{jf12}	JRC	N/A
Tg(UAS:CoChR-eGFP)^{jf44}	JRC	N/A
Tg(gfap:tdTomato)^{jf21}	JRC	N/A
Tg(gad1b:Gal4)^{jf49}	JRC	N/A
Tg(th-P2A-Gal4)	Chinese Academy of Sciences (Li et al., 2015)	N/A
Tg(UAS:GCaMP6s)	National Institute of Genetics	N/A
Tg(net:mTdTomato)	Cornell University (Farrar et al., 2018)	N/A
Tg(net:H2B-GCaMP6f)	Cornell University (Farrar et al., 2018)	N/A
Tg(gad1b:loxP-RFP-loxP-GFP)	Okazaki Institute for Integrative Bioscience (Satou et al., 2013)	N/A
Software and Algorithms		
MATLAB	MathWorks	https://www.mathworks.com
ImageJ	National Institutes of Health	https://imagej.nih.gov/ij/
C# (.NET Framework 3.5) (stimulus/behavior)	Microsoft	https://docs.microsoft.com/en-us/dotnet/csharp/
Cell segmentation and component detection	This study	https://github.com/mikarubi/segmentation
Chemicals, Peptides, and Recombinant Proteins		
Alpha-Bungarotoxin	Sigma-Aldrich	Cat# B137
Capsaicin	Sigma-Aldrich	Cat# M2028
Xestospongin C	Sigma-Aldrich	Cat# X2628
Prazosin hydrochloride	Sigma-Aldrich	Cat# P7791
Cyclazosin hydrochloride	Sigma-Aldrich	Cat# C247
Idazoxan hydrochloride	Sigma-Aldrich	Cat# I6138
Propranolol hydrochloride	Sigma-Aldrich	Cat# P0884
Methoxamine hydrochloride	Sigma-Aldrich	Cat# M6524

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Dextran, Texas Red	ThermoFisher Scientific	Cat# D3328
Dextran, Fluorescein	ThermoFisher Scientific	Cat# D3306
Metronidazole	MP Biomedicals	Cat# 02155710

LEAD CONTACT AND MATERIALS AVAILABILITY

Information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Misha B. Ahrens (ahrensm@janelia.hhmi.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Zebrafish

Larvae were reared at 28.5°C in 14:10 light-dark cycles (Westerfield, 2007). Zebrafish from 5 to 7 dpf were fed rotifers and used for experiments. All experiments complied with protocols approved by the Institutional Care and Use Committee at Janelia Research Campus (JRC). Zebrafish sex cannot be determined until ~3 weeks post-fertilization (Uchida et al., 2002), so experimental animals' sex was unknown.

Transgenesis

Transgenic zebrafish larvae were in *casper* or *nacre* background (White et al., 2008). *TgBAC(gad1b:loxP-RFP-loxP-GFP)* (Satou et al., 2013) was used without Cre-mediated recombination and called *Tg(gad1b:RFP)*. *Tg(elavl3:GCaMP6f)*^{jf1}, *Tg(elavl3:H2B-GCaMP6f)*^{jf7}, *Tg(elavl3:ReaChR-TagRFP-T)*^{jf10}, *Tg(elavl3:jRGECO1b)* lines are described (Chen et al., 2013; Dana et al., 2016; Dunn et al., 2016; Vladimirov et al., 2018). *Tg(gfap:jRGECO1b)*, *Tg(gfap:GCaMP6f)*, *Tg(gfap:CoChR-eGFP)*, *Tg(gfap:opto-α1AR-eYFP)*, *Tg(gfap:H2B-GCaMP6f)*, *Tg(dbh:KalTA4)*, *Tg(UAS:CoChR-eGFP)*, and *Tg(gfap:tdTomato)* lines were generated with Tol2 system (Urasaki, 2009), a known gfap promoter sequence (Bernardos and Raymond, 2006) and the following:

1. Construct containing dopamine-β-hydroxylase (dbh) promoter (dbh-GFP-pDEST-ISce); a gift from Thomas Look (Dana Faber Cancer Institute)
2. Construct containing transient receptor potential cation channel subfamily V member 1 (TRPV1) (or capsaicin receptor) from rat; a gift from David Prober (California Institute of Technology)
3. pEGFP-gfap (Intron1/5′/Exon1-zebrafish) (Addgene plasmid # 39761)
4. FCK-gene86-GFP (CoChR); a gift from Ed Boyden (Klapoetke et al., 2014)
5. pcDNA3.1-opto-α1AR-EYFP (Addgene plasmid # 20947)

Tg(gad1b:Gal4) was generated by CRISPR-mediated knock-in of Gal4 (Gal4FF) at gad1b locus as described (Kimura et al., 2014). The sgRNA sequence for gad1b is TGGAACTGCTCACCAGAAGG. 2XNRSE-tbait-hsp70-Gal4 (Kimura et al., 2014) was a gift from Shin-ichi Higashijima.

METHOD DETAILS

Preparation for fictive-behavior assays

All animals in fictive-behavior experiments were paralyzed and restrained. To induce paralysis, we bathed the fish in a 5 mg/mL solution of the nicotinic acetylcholine receptor antagonist α-bungarotoxin for 30 s, then transferred to fish facility water. Once fish no longer responded to tactile stimuli, we transferred them to a plastic Petri dish (for behavioral experiments) or a custom-fabricated sample holder (for light-sheet imaging). Paralyzed fish were then immersed in low-melting point agarose (Sigma) and oriented before it congealed.

Fictive behavior in virtual reality

For fictive swimming and virtual reality (Vladimirov et al., 2014), paralyzed zebrafish were mounted on an acrylic pedestal inside a square water-filled chamber with glass walls and a diffusive screen at the bottom for stimulus projection. Agarose was removed from the head and tail, and large-barrel glass pipettes (~60 μm inner diameter) were attached to the tail with gentle suction. The standard deviation of the left and right motor nerve signal (amplified by a MultiClamp 700B microelectrode amplifier) was calculated in a sliding window of 10 ms: this formed the 'swim signal' or 'swim power' used to quantify behavior.

In closed-loop one-dimensional environments, a stripe pattern (2.5 mm stripe width) was projected underneath the fish by a video projector (MP-CD1, Sony). This pattern moved forward at 2.86 mm/s drift speed when the fish was not swimming. When the fish

swam, the stripe accelerated backward at a rate proportional to swim power, i.e., [stimulus velocity] = [drift speed] – [motosensory gain] × [swim power]. This feedback mimics the visual feedback the animal would receive if it were freely swimming. When [swim power] = 0, forward velocity mimics the fish moving backward (as in a virtual water stream).

In open loop, the gain was zero, so the stripes moved forward at a fixed velocity, independently of whether the fish swam or not.

Brain imaging during fictive behavior used a custom light-sheet microscope, as described (Vladimirov et al., 2014). Briefly, two low-NA laser beams were scanned in the horizontal and vertical planes. The beams entered the chamber through the glass walls. One beam came from the side (relative to the head) and the other from the front of the fish and scanned the area between the eyes. The side laser was switched off at every sweep when pointing at the eye to avoid stimulating the retina. For every horizontal layer sweep, one image was acquired using a 16x/0.8 NA detection objective (Nikon), Nikon tube lens, and a camera (Orca Flash 4.0 v2, Hamamatsu). After each sweep, vertical position of the lasers and imaging objective was changed by 5 or 8 μm until a full brain volume was acquired (3 volumes/second). For GCaMP6s/f imaging, a blue laser (488 nm) was used for excitation with a green 525/50 nm detection filter (Semrock). Red light delivered visual stimulus. For jRGECO1b imaging, a green laser (561 nm) was used for excitation with a red 630/92 nm filter for detection (Semrock). For two-color imaging, 488 nm and 561 nm lasers were used simultaneously. An image splitter (W-View Gemini, Hamamatsu) separated a green and a red component (with a 525/50 nm detection filter (Semrock) for green emission and a 645/75 nm detection filter (Chroma) for red emission); red and green images were recorded on opposite halves of the same camera chip. Blue light delivered stimulus (animals showed optomotor response to either blue or red light from the video projector).

Futility-induced passivity behavioral analysis

Futility-induced passivity was defined by the inter-swim-interval. Paralyzed fish in closed loop in our VR setup typically swam in discrete bouts, about once per second (1.2 ± 0.05 s, n = 74). When switched to open loop, after swimming for ~20 s, fish paused for ~20 s in the passive state. About 50% of fish showed this behavior at 6-7 d.p.f.; fish that did not become passive in the open loop protocol were not included in subsequent analyses (50%). The active/passive state switching created a bimodal distribution of inter-swim intervals (Figure S1A), according to which we chose 5 s as the threshold defining passive epochs. An active epoch was defined as a chain of swim bouts separated by intervals under 5 s with more than 5 s between the start of the first swim bout and the end of the last bout. The two types of intervals segregated epochs well (1.19 ± 0.047 s inter-swim-interval in active epochs versus 24.4 ± 2.5 s in passive epochs, n = 74, Figure S1B). The average length of active epochs was 36.1 ± 4.2 s, and the fish swam 29 ± 6 times (n = 74) during each epoch.

DMD-based optogenetic stimulation

To target specific populations of cells for optogenetic stimulation, we built two photostimulation modules using digital micromirror devices (DMD) from commercial projector displays (P-300, AAXA; Lightcrafter discovery kit, Texas Instruments; see Optical Pathway for Optogenetic Stimulation). One photostimulation module was integrated to a standard epifluorescence microscope, and the other was integrated to our light sheet microscope. The optical design of our photostimulation modules is described (Zhu et al., 2012a). Briefly, the DMD was illuminated by a blue laser (488 nm in setup 1; 473 nm in setup 2). Patterned light reflected off the DMD was spatially filtered with a pinhole to select the 0^{th} diffractive order, then the relayed image of the DMD was expanded and projected onto the sample plane of the microscope. We wrote C#, Python, and Arduino software to control the lasers and create spatial patterns for photostimulation (available upon request).

Parameters for optogenetic stimulation: For *Tg(dbh:KalTA4);Tg(UAS:CoChR-eGFP)* fish on setup 1, the excitation intensity was 0.17 mW / mm^2 and the stimulation duration was 3 s of 10 ms pulses at 10 Hz. For *Tg(gfap:opto-α1AR-eYFP)* fish on setup 1, the excitation intensity was 0.42 mW / mm^2; the stimulation duration was 10 s of continuous illumination. For *Tg(gfap:CoChR-eGFP)* fish on setup 2, the stimulation intensity ranged from 0.23 to 0.46 mW / mm^2; the stimulation duration was 3 s of 10 ms pulses at 10 Hz.

Two-photon ablation of cells

For plasma ablation of individual radial astrocytic cells, we used a two-photon laser (Coherent, Chameleon Ultra II). Cells were selected based on their location between rhombomere 2 and rhombomere 6, along the midline of the dorsal hindbrain, and a genetic marker in *Tg(gfap:H2B-GCaMP6f)* fish. Ablation was done with high laser power (930 nm, 500 mW, measured after the objective) directed to a single point on each cell with short exposure time (1 – 4 ms) and 300–400 ms between ablations to prevent thermal damage. After successful ablation, an image stack showed previously bright, round nuclei appeared dim, irregular or fragmented. Cells without this signature were selected again for the next round. Each fish underwent 3–5 rounds of ablation. The experiment started 30-60 min after the last ablation.

Noradrenergic neurons in LC and norepinephrine cluster of the medulla oblongata (NE-MO) were similarly ablated in *Tg(th-P2A-Gal4); Tg(UAS:GCaMP6s)* fish and serotonergic neurons in the dorsal raphe nucleus (DRN) were ablated in *Tg(elavl3:H2B-GCaMP6f)*. In this line, DRN cells were identified by their anatomical location. In control experiments, similar numbers of neurons labeled by *Tg(elavl3:H2B-GCaMP6f)* in nearby brain regions were ablated with the same protocol.

Forebrain lesioning

Fish were embedded in 2% agarose and anesthetized in external solution with tricaine (0.2 mg/mL) at 4 dpf. An incision was made at the top of the forebrain, and a glass pipette (inner diameter ~20 μm) was inserted until it touched the brain. Negative pressure was applied through a tube connected to the glass pipette, and the forebrain was removed by suction. Fish were released from agarose twenty minutes later and returned to the incubator. Behavior experiments were done at 6 dpf, when the forebrain was still largely missing. Siblings were used as controls. Control fish went through the same procedure except that their forebrain was left intact. During testing, two fish of similar body size, forebrain-intact and forebrain-removed, were put in the same rig and recorded simultaneously.

TRPV1-mediated activation of radial astrocytes

For TRPV1-mediated activation (Chen et al., 2016) of radial astrocytes, we injected *gfap:TRPV1-T2A-eGFP* plasmid into embryos from *Tg(gfap:jRGECO1b)* transgenic fish at the 1–2 cell stage. TRPV1 protein expressed sparsely in radial astrocytes. Fish were screened at 3 dpf: larvae with strong, sparse GFP expression in the hindbrain were the 'TRPV1 (+)' group; those without detectable GFP expression were the 'TRPV1(-)' group. Both groups were screened again based on expression of jRGECO1b, and subdivided into four groups: 'TRPV1(+) jRGECO1b(+)'; 'TRPV1(+) jRGECO1b(-)'; 'TRPV1(-) jRGECO1b(+)'; and 'TRPV1(-) jRGECO1b(-)'. All fish were tested behaviorally; jRGECO1b-positive fish were imaged in a light-sheet microscope. For all experiments, fish were recorded for at least 0.5 h, and then capsaicin (Sigma; 5 μM final concentration in system water) or saline was delivered into the recording chamber, and recording continued for 1 to 2 h.

Calcium responses and behavior were quantified in two 10-min windows: from 11 min to 1 min before capsaicin ('before' treatment) and from 10 to 20 min after capsaicin ('after' treatment).

IP_3 receptor blockade

To partially block the calcium concentration increase in radial astrocytes, we inhibited IP_3 receptor-dependent calcium release by loading the inhibitor Xestospongin C (Sigma) (Gafni et al., 1997) near astrocytic cell bodies. Glass pipettes were pulled on a micropipette puller (P-1000, Sutter Instruments), and broken on a microforge (MF900, Narashige), to get a 10 μm-wide tip. This sharp glass pipette was loaded with 5 μL of a mixture of Xestospongin C (100 μM, in external solution) and a fluorescent dye (Dextran, Texas Red, 0.5%, or Dextran, Fluorescein, 0.5%, Thermofisher) to visualize the ejection. The pipette was then inserted through the skin of a paralyzed *Tg(gfap:GCaMP6f)* fish and advanced to the hindbrain ventricle, near astrocytic somata. We first assessed how long the fish spent in a passive state while in open loop for 30 min. Then, pulses (20 psi, 30 ms, inter-pulse-interval 30 s, 10 pulses) were delivered using a pressure ejection system (Picosprizter III, Parker Hannifin). Fish were then switched to the closed-loop condition for ~20 min to allow the drug to take effect and then back to open loop to record the time spent in passivity.

Analysis of fluorescence imaging data

Calcium indicators

Multiple calcium indicators - GCaMP6f, GCaMP6s, H2B-GCaMP6f, jRGECO1b - were used for different purposes: single color imaging, dual-color simultaneous two-population imaging, imaging with optogenetic stimulation. For whole-brain imaging analysis, we tested both the calcium indicators (jRGECO1b and GCaMP6f) in both neurons and radial astrocytes, and found similar relationships between calcium responses and behavior with both indicators.

Summary of Calcium Indicators for Each Experiment

Cell type	Calcium indicator	Figures
Neurons	GCaMP6f	Figures 2A–2D, 2F, 2H, 2J–2M, 6A, and S2A–S2E
	H2B-GCaMP6f	Figures 4E, S1K, and S5F
	jRGECO1b	Figures 5R, S5G, and S5H; tested for consistency
Radial astrocytes	GCaMP6f	Figures 2D, 2G, 2H–2M, 4H, 5I (bottom), 7K, S2C–S2E, S4I, and S6G
	jRGECO1b	Figures 2A–2C, 4G, 5B, 5D, 5E, 5F, 5H, 5I (top), 6C, 6H, 6K, 6L, 6N, 7H, S2A, S2F, S5A, S5C, S5H, and S6F
	H2B-GCaMP6f	Figures 4B–4D, 4F, and 4G
Noradrenergic neurons	GCaMP6s	Figure 6B, 6C, 6E, 6F, 7B–7I, and S6H

Baseline normalization

For all functional imaging videos, baseline fluorescence was estimated by smoothing each pixel's time series with a sliding-window percentile filter that estimated the 20th percentile of the data in 5-min windows, after subtracting the camera baseline (~100). ΔF/F was then calculated for each pixel's time series (after subtracting camera baseline) by subtracting the baseline from the raw time

series and dividing this difference by the baseline plus an offset of 10, to prevent division by baseline values very close to 0. The addition of an offset to the denominator of the ΔF/F calculation underestimates true ΔF/F but reduces the rate of artifacts.

Extraction of cells from voxel data

We developed a data-processing pipeline to extract populations of cell bodies and neuropil, then cluster these cell segments (called 'cells') into functionally relevant components (cells with similar activity patterns and reproducible anatomical structure), and finally to align and compare cells and components at specific behavioral points and across fish.

For each fish, our analyses began with ANTs 2.1.0 rigid-body registration of all volumes to the average volume (Avants et al., 2010). Then we created an intensity-based brain mask and fully divided the brain into about 1000 spatially contiguous three-dimensional blocks, which overlapped slightly to capture cells on the borders.

We used the Spark cluster computing framework (Zaharia et al., 2010) to parallelize cell-detection across blocks. Our algorithm was based on constrained non-negative matrix factorization. For n voxels, t time points, and c cells, we factorized

$$V_{(n \times t)} \approx W_{(n \times c)} H_{(c \times t)} + X_{(n \times 1)} I_{(1 \times t)} \quad (1)$$

where V is the full spatiotemporal fluorescence matrix for each block, W and H are, respectively, the spatial location and time series of segmented cells, and X and I are a rank-1 spatiotemporal model of the background signal. We set the block size under the assumption that each block would contain 100 tightly packed 6 μm diameter spheres. Thus, we set the number of cells in each block to be 100, and then gradually reduced this number (through iterative multiplication by 0.95) until the system in Equation 1 was not rank-deficient and factorization converged.

We combined local intensity peaks and local correlation coefficients to initialize W (and used a constant value to initialize X); and approximately solved Equation 1 using alternating least-squares (Berry et al., 2007), with at least 10 iterations, and at most 100 iterations or until numerical convergence (defined as a change in the value of the residuals < 0.001).

We imposed spatial and temporal constraints to regularize this factorization, including mean-amplitude normalization of each cell; hard spatial constraints: retaining the largest connected segment within a static 12 μm diameter sphere (centered on a local-intensity peak); and soft sparseness constraints: a sparse projection of the spatial footprint for each cell using the algorithm of Hoyer (Hoyer, 2004), with sparseness,

$$\text{sparseness}(W_i) = \frac{\sqrt{n} - \sum_j |W_{ij}| \Big/ \sqrt{\sum_j |W_i|^2}}{\sqrt{n} - 1} \quad (2)$$

set to be at least equal to the spatial sparseness of a binary 6 μm diameter sphere (these constraints were not applied to the model of the background signal)

The resulting cell representation of whole-brain activity accurately reproduced the raw voxel time series and had increased signal-to-noise ratio due to demixing and denoising. Regions of neuropil and astrocytic processes were factorized similarly to cell bodies, also yielding small patches of voxels with consistent signals.

Clustering cells into components

We clustered segmented cells into functionally relevant components in two steps. First, we used non-negative matrix factorization to reduce the time series for all segmented cells in the brain into low-dimensional components. Specifically, for k cells, t time points, and $m = 60$ components, we factorized

$$H_{(k \times t)} \approx S_{(k \times m)} T_{(m \times t)} \quad (3)$$

where H are time series for all segmented cells in the brain, while S and T are the spatial and temporal low-dimensional components. These components need not be spatially localized, so we substituted local intensity-based initialization for nonnegative double SVD method (Boutsidis and Gallopoulos, 2008) and imposed no spatial constraints on this factorization.

Second, we detected components in this reduced representation. We deemed a component to be functionally relevant when it contained cells with similar activity patterns and showed robust and reproducible anatomical structure across fish (Figures 2D and 2E; Video S2). In contrast, noise components lacked clear spatial structure or were not reproducible across fish.

To evaluate reproducibility, we constructed an average brain template, independently for neurons and radial astrocytes, using the nonlinear registration tool antsMultivariateTemplateConstruction2 (Avants et al., 2010). We registered the spatial component of each fish to this template, and did semi-automated clustering across all components of all fish. We clustered spatial components by binning adjacent voxels of each component, computing the correlation coefficient between all components, thresholding the correlation matrix, and clustering the thresholded correlation matrix using the clique percolation method (Palla et al., 2005), an algorithm that can flexibly assign each component to more than one cluster or to no clusters. We manually fine-tuned this initial clustering by excluding components from clusters if the average spatial correlation of the component with the cluster was too low, and including the component if the average correlation of the component to all components within the cluster was sufficiently high. This step yielded 10-20 neuronal and astrocytic clusters. All components within each cluster were averaged, smoothed, and warped to individual fish space. Figures 2D and 2E shows the 10 group-average components with highest p values (see below), while Video S2 shows all group-average components created in this way for neurons and radial astrocytes.

To avoid misalignment of group-average components in individual fish, we used an additional nonnegative matrix factorization step, which aligned the components more tightly to each fish. Specifically, we refactorized H (Equation 3), with S initialized by and constrained to be within the binary mask of each group-average component. This added step allowed us to combine the consistency of components across multiple fish with the individual spatial variation of components within each fish. Finally, we averaged the activity of cells within the spatial mask of each component in individual fish, and compared this average across active and passive states.

Analysis of component and cell activity around passivity onset

We related neuronal and astrocytic component and cell dynamics to behavior by upsampling all calcium signals to 10 Hz, and registering them to passivity onset, and to the midpoints of bouts preceding this onset. As timings of individual bout sequences varied (relative to passivity onset) across trials, we registered bouts by averaging the relative times of bout sequences across trials and fish and linearly interpolating between individual-trial and average bout-sequence times.

This approach allowed us to study average activity of each component and cell relative to passivity onset (Figures 2F–2M; Video S3). We estimated the increase in calcium during and after passivity onset by comparing a sliding 1 s window of $(\Delta F/F)$ to the baseline, defined as the average over 10 to 5 s before passivity onset. We computed the significance of any increases in calcium by one-tailed Wilcoxon signed-rank test across trials. At this exploratory stage, we sought to identify candidate components and cells associated with passivity onset by computing $-\log_{10}(\text{p value})$, where p values from the Wilcoxon rank-sum were averaged between 1 s to 4 s after passivity onset.

We further analyzed the temporal profile of activity of neurons and radial astrocytes close to passivity onset at the single-cell level, by averaging activity of each segmented cell within a one-second window centered on the peak of each bout, and fitting an exponential function to the activity averaged over the last five bouts before passivity onset (Figure 6A). We interpreted the best-fit exponent of the cell as the rate at which single-cell fluorescence increased before the fish became passive.

Confocal microscopy

Confocal imaging was done using a Zeiss LSM 710, a Zeiss LSM 880 upright confocal, and a Zeiss LSM 800.

Single-cell radial astrocyte tracing

Tg(gfap:jRGECO1b) or *Tg(gad1b:RFP)* fish at 6 dpf were embedded dorsal side up in 2% agarose in a 35 mm Petri dish. The head was exposed by cutting away the agarose above the hindbrain. Fish were then imaged with a two-photon microscope at a wavelength of 1030 nm in the dorsal anterior hindbrain adjacent to the Lateral Medulla Oblongata (L-MO), at the level of rhombomeres 2-3, the approximate location of the rostral cluster of radial astrocytes. Single radial astrocytes were electroporated with 20% Texas Red Dextran or fluorescein Dextran, 3,000 MW, anionic, lysine fixable (Molecular Probes) as described (Koyama et al., 2011). For electroporation, we delivered 150 square pulses of 1 ms duration and 30 V at a frequency of 20 Hz using glass electrodes with an impedance of 240 MΩ. Then fish were released from agarose and transferred to the incubator. After overnight fixation in a 4% solution of paraformaldehyde in phosphate-buffered saline (PBS) pH 7.4 at 4°C, fish were dissected to expose the brain and imaged from the dorsal side on a Zeiss LSM 880 confocal microscope using a 25x objective. Confocal stacks were analyzed with FIJI (ImageJ).

In situ hybridization

Zebrafish larvae at 6 dpf were euthanized with an overdose of MS-222 (tricaine) and fixed overnight in 4% PFA-PBS, pH 7.4 at 4°C. Fish were dissected to expose the brain and processed for whole mount *in situ* hybridization (WISH) using the RNAScope technology by manufacturer protocol (Gross-Thebing et al., 2014). The adra1Bb custom probe was designed by the manufacturer and used on transgenic *Tg(gfap:GCaMP6f)* fish. Fish were then mounted in Vectashield (Vector Laboratories, Burlingame) and imaged on a Zeiss 800 upright confocal microscope with a 25x water immersion objective.

Immunohistochemistry with expansion microscopy

Tissue preparation

Fish brains were dissected out in PBS and fixed in 4% PFA-PBS, pH 7.4 overnight at 4°C, then rinsed well with PBS before antibody labeling.

Antibody staining

Brains were incubated in blocking buffer (10% goat serum, 0.8% Triton X-100 in PBS) overnight at 4°C. Brains were then incubated with primary antibodies chicken anti-GFP (Aves, GFP-1020) and rabbit anti-dsRed (Clontech, 632496) diluted 1:500 in blocking buffer, for 2 days at 4°C, followed by washing in wash buffer (0.3% Tween 20 in PBS) 5x 30 min at room temperature. Brains were incubated in fluorescent secondary antibodies goat anti-chicken Alexa Fluor 488 (Invitrogen, A11039) and goat anti-rabbit Alexa Fluor 594 (Invitrogen, A11037) diluted 1:500 in blocking buffer, for 2 days at 4°C, followed by washing in wash buffer 5x 30mins at room temperature. Brains were then washed and stored in PBS. All antibody staining steps included shaking.

Expansion

Expansion was done as described (Tillberg et al., 2016) (see also Freifeld et al., 2017). AcX was resuspended at 10mg/mL in anhydrous DMSO, aliquoted, and stored desiccated at –20°C for up to several months. AcX aliquots were diluted 1:100 in PBS to produce anchoring solution just before use. Specimens were incubated with at least 100-fold volume excess of anchoring solution, at room temperature overnight, protected from light, without shaking.

Gelation solution (8.6% (w/v) sodium acrylate, 2.5% (w/v) acrylamide, 0.15% (w/v) bisacrylamide, 2M sodium chloride, 1xPBS, 0.01% (w/v) 4HT, 0.2% (w/v) TEMED, and 0.2% (w/v) APS in milliQ water) was prepared as follows: all components except 4HT, TEMED, and APS were mixed to produce StockX, which was aliquoted and stored at −20°C for up to several months. Aliquots were thawed and chilled to 4°C, and then 4HT, TEMED, and APS were added from concentrated stocks (0.5% (w/v) 4HT, 10% (v/v) TEMED, 10% (w/v) APS) just before use. Specimens were incubated with at least 100-fold volume excess gelation solution 2x 10min at 4°C.

During the incubation in gelation solution, chambers were prepared by applying adhesive silicone gaskets (ThermoFisher Scientific, P24743) to plain glass slides and coating the glass bottom of each well with 1 μL poly-lysine coating solution (poly-lysine solution (Ted Pella 18026) with photo-flo (EMS 74257) added up to 0.2% (v/v)), leaving at least 5 min for poly-lysine coating solution to dry onto glass.

Following incubation in gelation solution, specimens were transferred to the middle of chambers, where they adhered to the poly-lysine treated surface, leaving at least 2 mm between each specimen and the chamber wall. Excess gelation solution was removed and replaced with 36 μL gelation solution for each 9 mm diameter chamber. Chambers were sealed with cover glass, which forms a seal with the silicone gasket under gentle pressure. Care was taken to avoid bubbles within 2 mm of specimens. Chambers were then held at 37°C for 2 h to gel and cure specimens. Chambers were moved to 37°C within 30 min of adding APS to the solution to prevent premature gelation.

After gelation, chambers were brought back to room temperature and disassembled. Top cover glass and silicone gaskets were carefully removed, leaving gels adhered to glass slides. Gels were trimmed using a razor blade to a right trapezoidal shape to aid in post-expansion orientation of the specimen. Several microliters of digestion buffer (0.5% (w/v) Triton X-100, 500 mM sodium chloride, 1 mM EDTA, 50 mM Tris pH 8) were applied around the edge of each gel to de-adhere them from the glass surface. Proteinase K (New England Biolabs, P8107S) was added 1:100 to digestion buffer to produce digestion solution. Gels were incubated in at least a 100-fold volume excess of digestion solution overnight, at room temperature, with shaking. Following digestion, gels were washed several times in PBS and kept at 4°C until further use.

Expansion and imaging

Gels were expanded by washing in water and imaged on a Zeiss Z1 light sheet microscope with standard settings (Figure 3) or a Zeiss LSM 880 confocal microscope (Figure S3). Adjacent volumes were stitched using the Zeiss ZEN software. In Figures 3 and S3, stitching artifacts arising from non-ideal software stitching are indicated with arrowheads and dashed lines.

Materials Used Specifically for Expansion Microscopy

Name	Formal name	Supplier	Cat #
AcX	Acryloyl-X, SE, 6-((acryloyl)amino) hexanoic Acid, Succinimidyl Ester	ThermoFisher Scientific	A20770
Sodium acrylate	Sodium acrylate	Sigma	408220
Acrylamide	Acrylamide	Sigma	A9099
Bisacrylamide	N,N′- Methylenebisacrylamide	Sigma	M7279
4HT	4-Hydroxy-TEMPO	Sigma	176141
TEMED	N,N,N′,N′- Tetramethylethylenediamine	Sigma	T7024
APS	Ammonium persulfate	Sigma	A3678

RNA sequencing

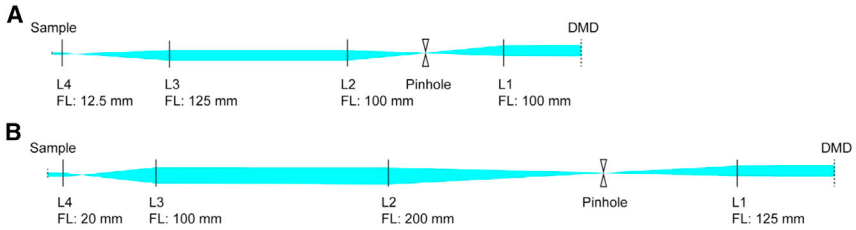
Cell isolation, library preparation, and sequencing

Neurons and radial astrocytes were isolated using methods of Hempel et al. (2007) with minor modifications. Fish were euthanized with 1:30 dilution MS-222/Tricaine in external solution, then placed on ice. Brains were dissected in 1:30 dilution of MS-222/Tricaine in external solution, then digested with Liberase-DH (1 mg/mL; Roche) in external solution for about 1 h. Digested brains were washed three times in external solution with 1:30 dilution of MS-222/Tricaine. Digested brains were triturated as described (Hempel et al., 2007) in external solution containing Tricaine and 2% fetal bovine serum. Cells were isolated by hand. About 300 cells were collected for each sample and lysed in 50 μl PicoPure extraction buffer. RNA was extracted and libraries were prepared as described (Cembrowski et al., 2016). Briefly, cDNA was prepared from DNase-treated RNA using Ovation RNA-seq v2 kit (NuGEN). Barcoded libraries were prepared with Ovation Rapid library kit (NuGEN) and sequenced on a HiSeq 2500 to a depth of about 40 M reads per sample with single-end 100 bp reads.

Gene expression estimation

Sequencing adapters were trimmed from the reads using cutadapt (https://doi.org/10.14806/ej.17.1.200). Trimmed reads were aligned to danRer10 Refgene transcriptome (UCSC). Gene expression estimates were reported as transcripts per million.

Optical Pathway for Optogenetic Stimulation



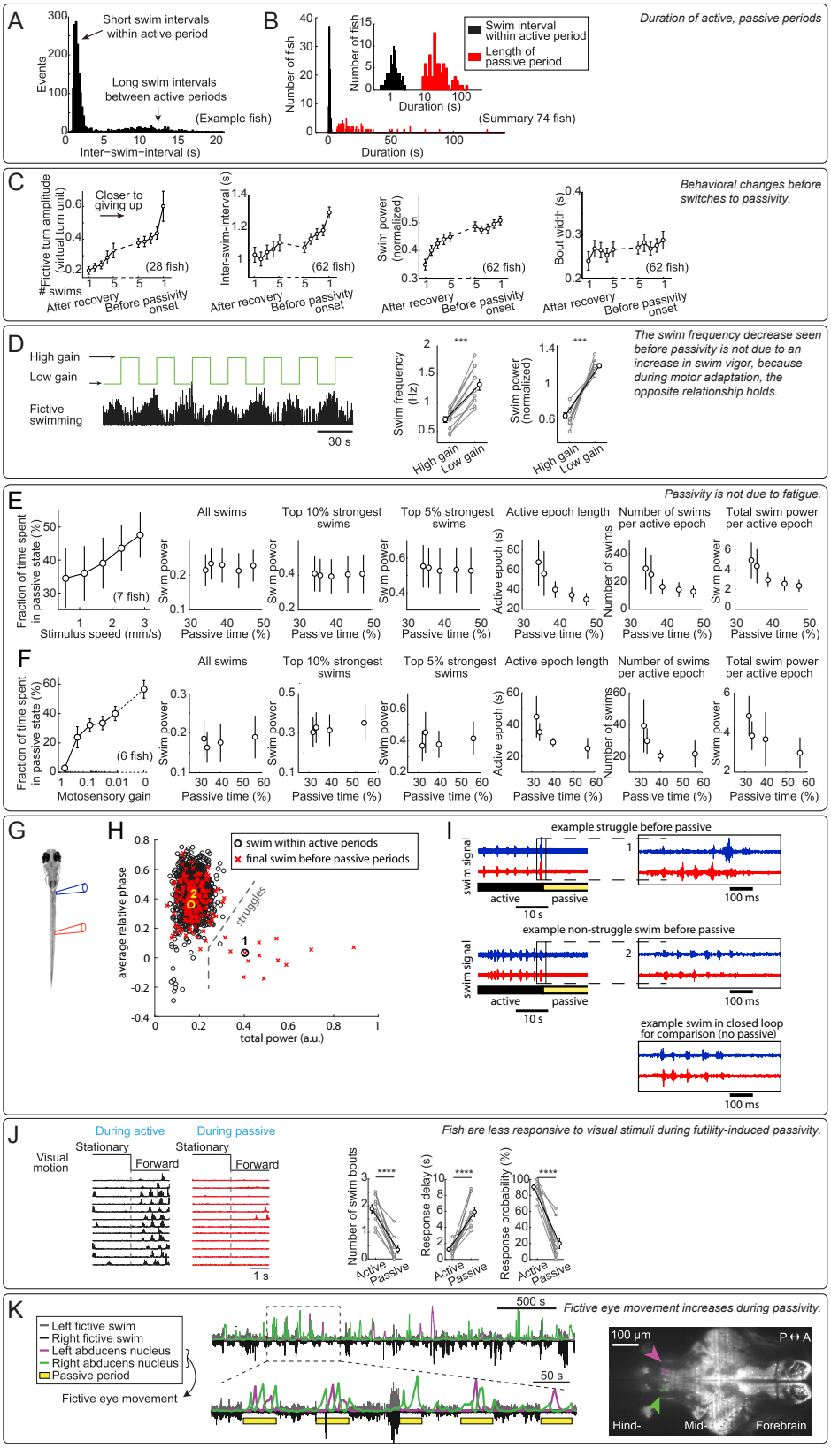
Optics in the two DMD photostimulation modules. The chief difference between these designs is demagnification of DMD image at focal plane of L4 (objective lens).

(A) Setup 1, photostimulation module in light sheet microscope. This module was used for all experiments that combined optogenetics with calcium imaging. As this module is designed for one photostimulation wavelength, lenses F1-F3 in panel A are singlets (Thorlabs). Final demagnification of DMD chip is 0.1X.

(B) Setup 2, Photostimulation module used in epifluorescence microscope. This module was used for all optogenetic experiments without calcium imaging. As this module is designed for two photostimulation wavelengths, lenses F1 - F3 in Panel B are achromatic doublets (Thorlabs). The final demagnification is 0.32X.

DATA AND CODE AVAILABILITY

The code for cell segmentation and component detection is available at https://github.com/mikarubi/segmentation.



(legend on next page)

Figure S1. Higher-Order Behavioral Effects of Futility-Induced Passivity, Related to Figure 1

(A and B) Distribution of inter-swim intervals in open-loop. Intervals followed a bimodal distribution: short swim intervals occurred within the active period when fish swam once every 1-2 s, and long swim intervals occurred between active periods when the fish stopped swimming (defined as the 'passive period'). (A) Distribution of all inter-swim-intervals from an example fish. (B) Distribution of average inter-swim-intervals during classified active periods (black bars), and average duration of passive periods (red bars) across fish.

(C) Over the course of an active period in open-loop, the animals: 1. made larger turns, and turning amplitude was largest just before passivity onset; 2. swam less frequently, and the longest interval between swims happened just before passivity onset; 3. swam more vigorously, and swimming was on average strongest just before passivity onset; 4. but the duration (width) of swim bouts was not strongly modulated. Error bars, SEM.

(D) Decrease in swim frequency (seen before passivity in C) is not due to an increase in swim power, control for Figure 1I. (*Left*) Motor adaptation assay; low motosensory gain in a closed-loop environment led to more powerful swimming. (*Right*) During low motosensory gain, swim frequency and swim power were both elevated compared to during high motosensory gain. Therefore, the decrease in swim frequency observed as the animal approaches passivity (Figure 1I) was not simply due to a concurrent increase in swim power. (Two-tailed paired t test, ***$p < 0.001$, $p = 8.9 \times 10^{-6}$ for swim frequency, $p = 2.7 \times 10^{-6}$ for normalized swim power. Error bars, SEM)

(E and F) Effects of stimulus velocity and motosensory gain on futility-induced passivity imply that passivity was unlikely to be caused by fatigue. More passivity was caused by (E) faster forward-moving visual stimulus in open loop, or (F) lower motosensory gain in closed loop. Under those conditions, increased occurrence of passivity was not due to more vigorous or more frequent swimming because: similar swim power of individual swims was observed for different velocities or gains, as well as of the top 10% strongest swims and the top 5% strongest swims. On the contrary, more passivity can be associated with smaller amounts of total swim output: shorter active periods, fewer swim attempts, and less total swim power. Together, these results imply that passivity was not caused by fatigue. Error bars, SEM.

(G–I) During transitions from active periods to passive periods, fish occasionally struggle. (G) Experimental setup showing placement of an electrode pair, recording electrical signals from the same side but different segments of the tail (Liao and Fetcho, 2008). (H) Struggles can be identified by the relative phase difference between anterior and posterior electrodes, and power per burst. Final swims before passivity can be non-struggle swims or struggles. On average 13 ± 9% of switches are preceded by a struggle (identified by burst power, n = 5 fish, mean ± SEM). (I) Example of switches preceded by a struggle (top), non-struggle swim (middle), and for comparison, closed-loop swimming (bottom).

(J) Fish were less responsive to visual stimuli during the passive periods. In an assay where, first, passivity was elicited by the usual paradigm, then a static stimulus was presented (2 s), and then pulse of visual motion was presented (2 s), the animals responded far less to the motion pulse if they were in the passive behavioral state. (Two-tailed paired t test, ****$p < 0.0001$, $p = 2.6 \times 10^{-6}$ for swim bout number change, $p = 3.8 \times 10^{-5}$ for response delay, $p = 4.2 \times 10^{-7}$ for response probability. Error bars, SEM) Same data as in Figure 1K.

(K) During the passive periods, calcium activity was seen in neurons in the abducens nucleus (arrows) in animals expressing nuclear-localized GCaMP6f (*Tg(elavl3:H2B-GCaMP6f)*) imaged with a light-sheet microscope. Neurons in the abducens nucleus are known to drive eye movement. Thus, increased activity in this area, alternating between left and right, implies that animals were attempting to move their eyes during the passive periods.

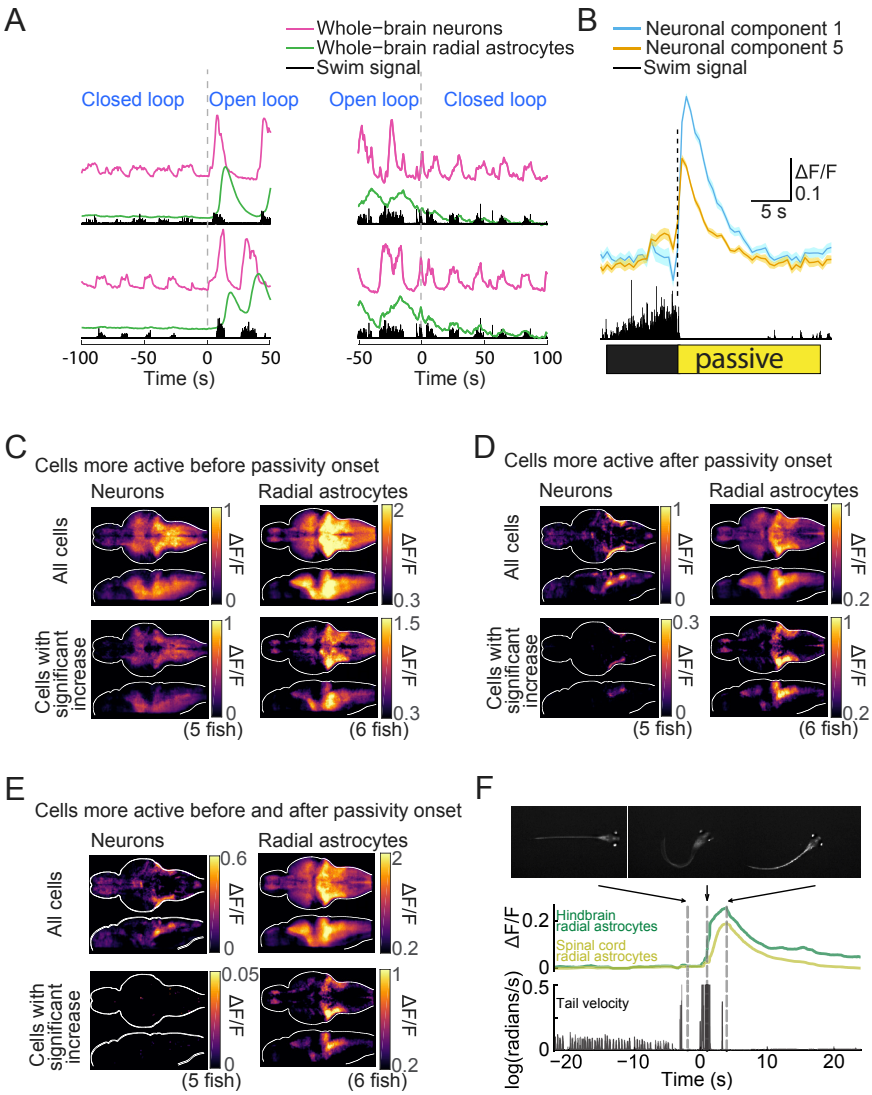


Figure S2. Whole-Brain Activity Elevation before and after Passivity Onset in Both Paralyzed and Non-paralyzed Fish and Passivity in Non-paralyzed Fish, Related to Figure 2

(A) Neuronal and astrocytic activity near the time of open-loop onset (left), and closed-loop onset (right). Signals are averaged across all the neurons or all the radial astrocytes in the brain, and two switches from each type are shown (top and bottom). Neurons showed stronger activity fluctuations during open-loop than during closed-loop, but showed clear activity patterns in both scenarios. In contrast, astrocytic activity fluctuations (calcium) were small before open-loop, but increased before and during passivity. Signals are z-scored.

(B) Average activity of neuronal component 1 and component 5, from the same fish in Figure 2H, triggered by passivity onset. Each trace is the average of 42 switches from active to passive states. Shading represents SEM.

(C–E) Whole-brain neuronal and astrocytic maps for cells showing activity elevation in windows before, after, or both before and after passivity onset (average over 5 fish for neuronal imaging, 6 for astrocytic imaging). Average ΔF/F signals from cells statistically increasing their activity relative to baseline (Methods),

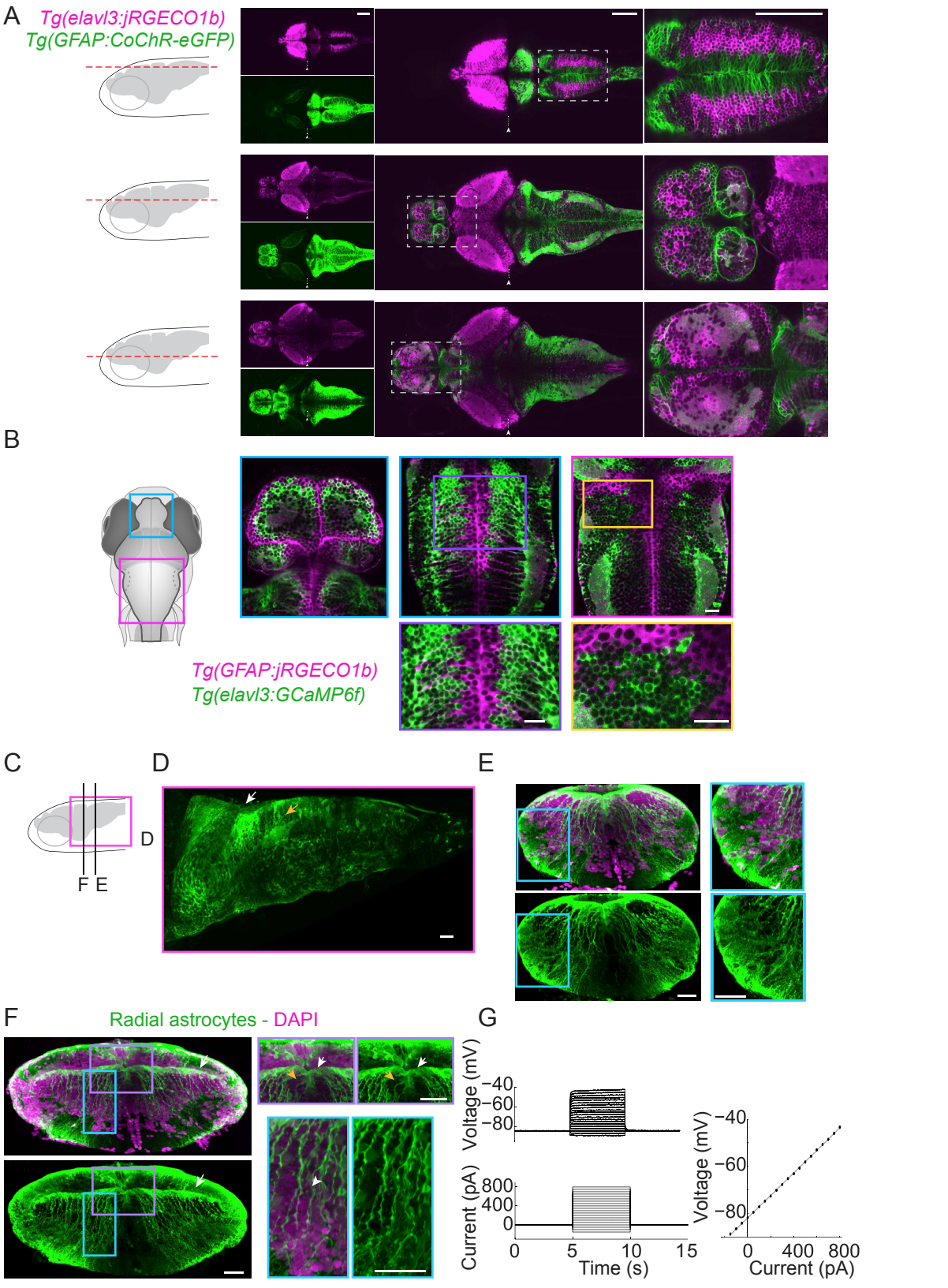
(C) before passivity onset (−1 to 0 s).

(D) after passivity onset (+1 to 4 s).

(E) both before and after passivity onset (−1 to 0 s and +1 to 4 s).

For each panel, top rows show average activity from all cells, and bottom rows show average activity only from cells with significant activity elevation.

(F) A zebrafish, expressing GCaMP6f in radial astrocytes, was placed in agarose. The agarose was slowly coagulating, but never fully prevented tail motion. The fish attempted to swim and the viscous agarose prevented the movement, qualitatively mimicking a reduction in motosensory gain. The animals displayed similar passivity as in the virtual reality assay. Calcium levels followed a similar pattern as was seen in futility-induced passivity in virtual reality. Calcium activity also increased in the spinal cord, but with a delay.



(legend on next page)

Figure S3. Anatomical and Electrophysiological Properties of Radial Astrocytes Projecting to the L-MO, Related to Figure 3

(A) Distinct expression patterns of CoChR-eGFP driven by the *gfap* promoter and jRGECO1b driven by the *elavl3* promoter suggests that the *gfap*-driven expression was largely restricted to radial astrocytes. Three columns represent three sections arranged from dorsal to ventral across the brain, indicated by the horizontal dashed lines in the schematic at the top of each column. *Row 1*, separated channels for CoChR-eGFP and jRGECO1b (the green channel was overexposed in order to visualize the low-level fluorescence in the tectum); *Row 2*, merged view of those two channels; *Row 3*, zoomed-in view of the region outlined by dashed lines in *Row 2*. Arrowheads indicate microscopy software stitching artifacts. Scale bars, 100 μm.

(B) Distinct expression patterns of jRGECO1b driven by the *gfap* promoter and GCaMP6f driven by the *elavl3* promoter, the combination being used for simultaneous imaging of radial astrocytes and neurons. Scale bars, 20 μm.

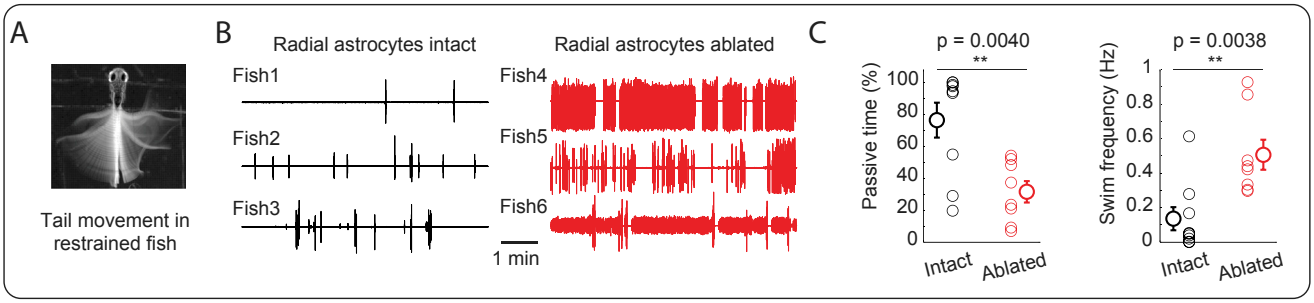
(C) Schematic side view of the fish brain showing the position of the panel (D) and the cross sections in (E) and (F).

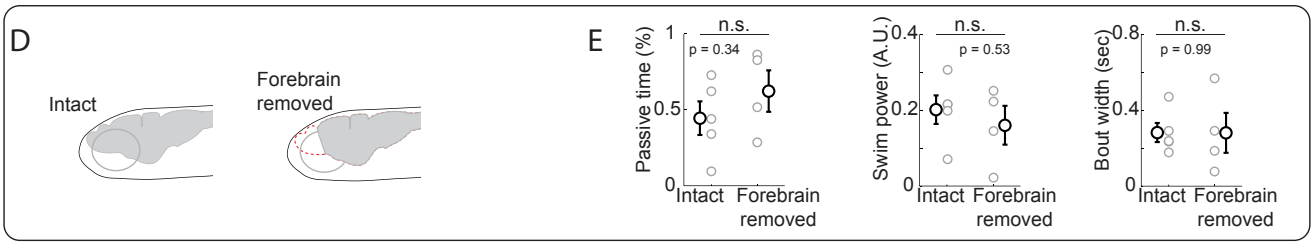
(D) Side view of a section of the hindbrain showing astrocytic cell bodies and processes (white arrow) and the lateral hindbrain neuropil (yellow arrow). Scale bars, 100 μm.

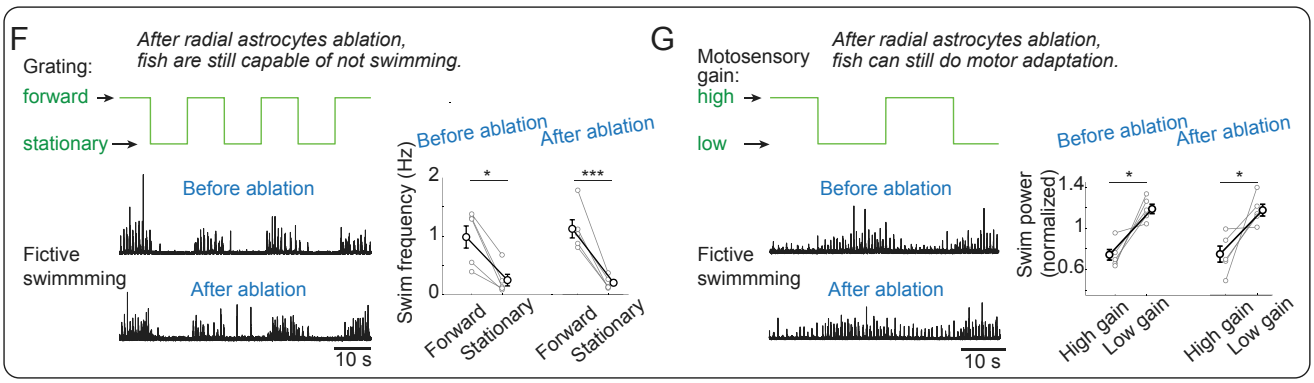
(E) Cross section from a level of the hindbrain that contains the lateral hindbrain (L-MO) (the cyan box inset contains the L-MO). Scale bars, 100 μm.

(F) Cross section at the level of the cerebellum showing the midbrain-hindbrain border (MHB) (arrow). *Top-Left,* Astrocytic cells are shown in green; DAPI staining labels nuclei of all cells (neurons and radial astrocytes). *Bottom-Left,* astrocytic cells only. *Purple box inset*, astrocytic cell bodies (white arrow) can be seen along the dorsal midline and the MHB. Neurons are dark without the GFAP label (yellow arrow). *Cyan box inset,* long astrocytic processes extend dorso-ventrally and terminate in the lateral and ventral gliopil. Scale bars, 100 μm.

(G) Patch-clamp recording of a GCaMP6f-positive cell from a *Tg(gfap:GCaMP6f)* fish shows electrophysiological properties of radial astrocytes: (1) a linear voltage change in response to injected current, (2) more negative resting membrane potential than typical neuronal membrane (−84 mV), consistent with known recordings of radial astrocytes in zebrafish (Li et al., 2015). Error bars, SEM.







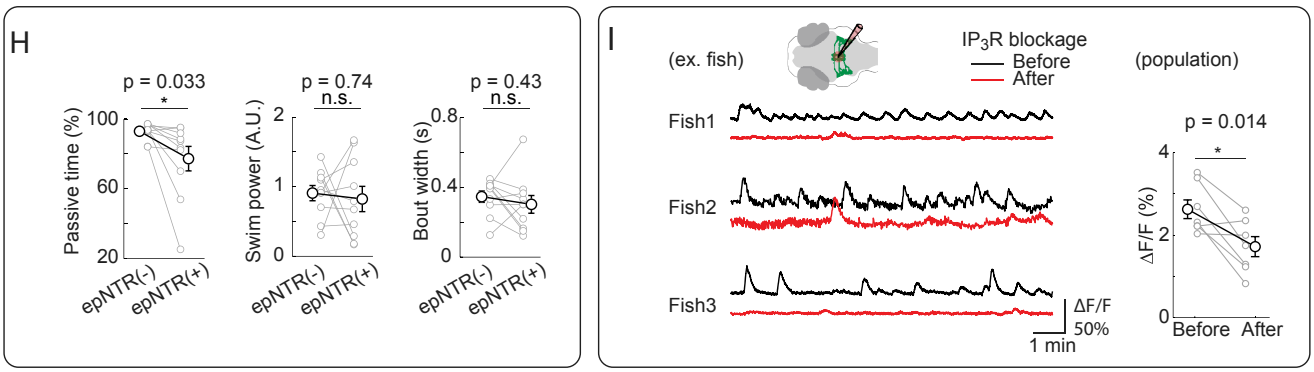


Figure S4. Details of Radial Astrocytes Inactivation, Related to Figure 4

(A–C) Astrocytic ablation in non-paralyzed zebrafish also reduces futility-induced passivity. (A) Fish were embedded in agarose, which was cut away around the tail, eyes and mouth. A time lapse image is shown during vigorous tail movement. (B) Non-paralyzed fish performed futility-induced passivity when head-embedded in agarose. With radial astrocytes intact, swimming occurred rarely. This may be due to the lack of visual feedback during tail movement, as well as the perception of pressure on the upper section of the trunk and head during tail movement. After laser ablation of radial astrocytes, the animals swam much more often. (C) Population data, showing a decrease in time spent passive and an increase in average swim frequency after astrocytic ablation (two-tailed non-paired t test, $^{**}p < 0.01$, $p = 0.0040$ for passive time, $p = 0.0038$ for swim frequency). Error bars, SEM.

(D and E) Removing the forebrain doesn't decrease passivity during open-loop. Removing the forebrain (D, Methods) did not reduce passivity during open-loop (E), showing that non-specific brain damage, of neurons or radial astrocytes, was not the reason for the reduction in passivity. Fictive swimming behavior was similar between the forebrain intact and forebrain-removed siblings (two-tailed non-paired t test). Error bars, SEM.

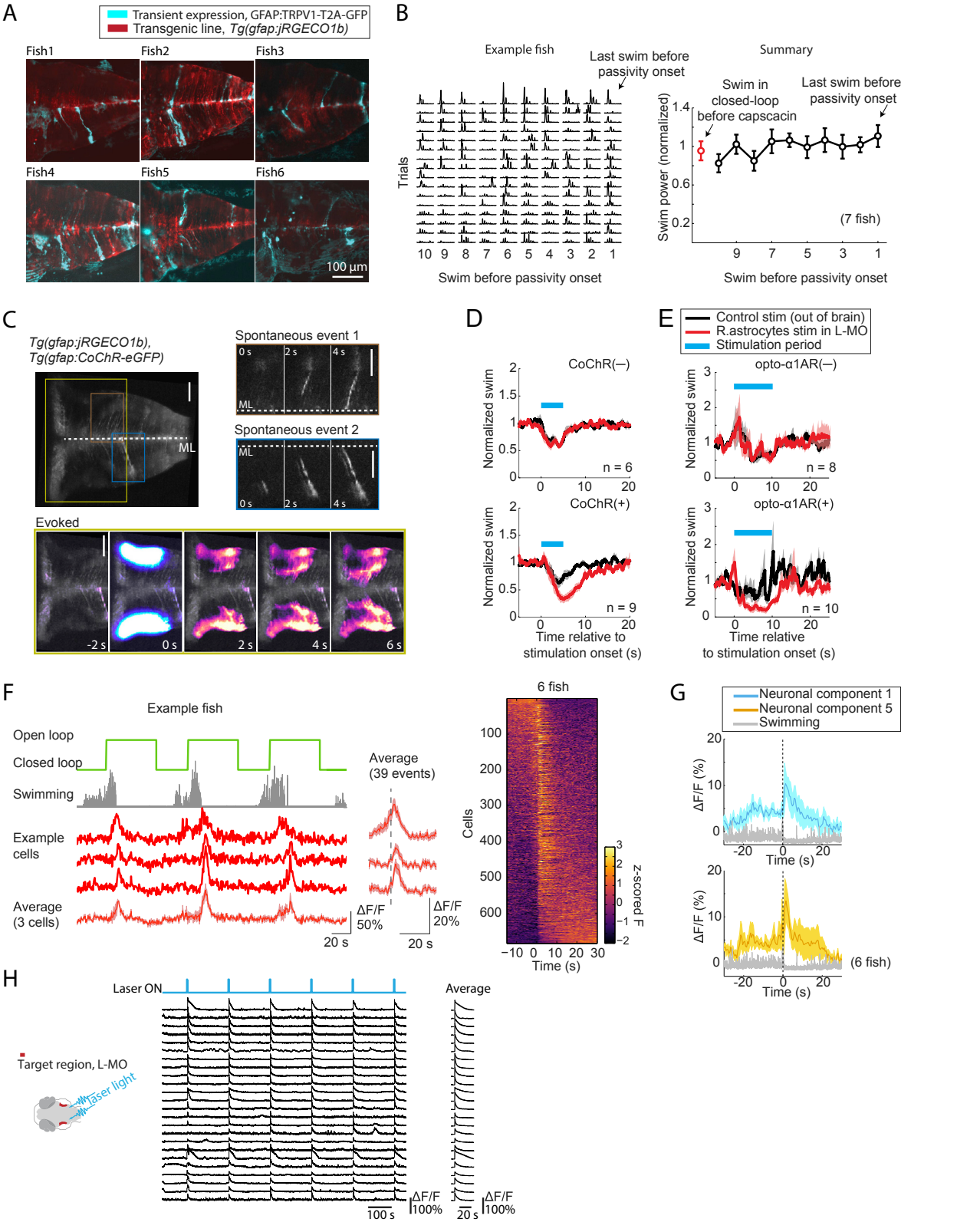
(F and G) Animals with radial astrocytes ablated can still perform motor adaptation and are still capable of not swimming. (F) When presented with forward visual motion alternating with a stationary stimulus, animals swam little during the stationary periods, both before and after astrocytic ablation, showing that ablation of radial astrocytes does not affect the ability of the animals to stop swimming. (G) Animals performed motor adaptation in a closed-loop environment, before and

(legend continued on next page)

after astrocytic ablation, showing that ablation of radial astrocytes does not affect the ability of the animals to modulate swim vigor. (Two-tailed paired t test, *$p < 0.05$, ***$p < 0.001$, $p = 0.014$ for before ablation, $p = 9.8 \times 10^{-4}$ for after ablation in (F); $p = 0.013$ for before ablation, $p = 0.042$ for after ablation in (G)). Error bars, SEM.

(H) Chemogenetic injury to radial astrocytes leads to a decrease in passive periods but leaves other swim behavior unchanged. The addition of a low concentration of the prodrug metronidazole (10 mM, incubation 8 h) (Tabor et al., 2014) caused intracellular damage to radial astrocytes in *Tg(gfap:epNTR)* fish that expressed nitroreductase in astrocytic cells. (Two-tailed paired t test, *$p < 0.05$, $p = 0.033$ for passive time, $p = 0.74$ for swim power, $p = 0.43$ for bout width). Error bars, SEM.

(I) Blocking inositol 1,4,5-trisphosphate receptor (IP_3R) reduces calcium activity in L-MO astrocytic cells during open loop. *Left,* calcium signals in astrocytic processes in the L-MO before (black) and after (red) injection of the IP_3R inhibitor, showing that calcium signals were reduced by the blockade of IP_3R. *Right,* population data showing that the reduction in calcium signal was consistent across animals. (Two-tailed paired t test, *$p < 0.05$, $p = 0.014$). Error bars, SEM.



(legend on next page)

Figure S5. Details of Chemogenetic and Optogenetic Activation of Radial Astrocytes, Related to Figure 5

(A) Sparseness of expression of TRPV1. Maximum intensity projections along the dorsal-ventral axis of hindbrain volumes acquired with a light sheet microscope show that TRPV1 (Chen et al., 2016) channels (introduced into the fish by injection of a *gfap:TRPV1-T2A-GFP*, plasmid into *Tg(gfap:jRGECO1b)* transgenic embryos) were sparsely expressed in radial astrocytes. Expression was seen in about 10-50 astrocytic cells per animal, but opening of the TRPV1 channel with capsaicin activated a much wider population of glial cells (at least hundreds, likely through communication between radial astrocytes). Expressing cells showed the morphology of hindbrain radial astrocytes, with somata along the midline and processes extending laterally and then ramifying. (Note, the images show some contamination from skin autofluorescence. Also note, the jRGECO1b expression was more patchy than usual, likely because the co-injected *Tol2* transposase removed the *jRGECO1b* transgene from the genome of some cells.)

(B) Artificially activating radial astrocytes induces passivity without increasing swim vigor, consistent with a role for radial astrocytes in specifically suppressing swimming. Passivity was caused in closed-loop by applying capsaicin to fish that expressed TRPV1 (Chen et al., 2016) in radial astrocytes. In the swim bouts just prior to TRPV1-mediated passivity, swim vigor did not increase as it did during futility-induced passivity (Figure S1C) – rather, swim vigor stayed at a similar level as swims in closed-loop without capsaicin, showing that radial astrocytes caused passivity independently of increases in swim vigor. In example fish (left panel), each row represents the final 10 swims of an active period before the fish switched to a passive state. Error bars, SEM.

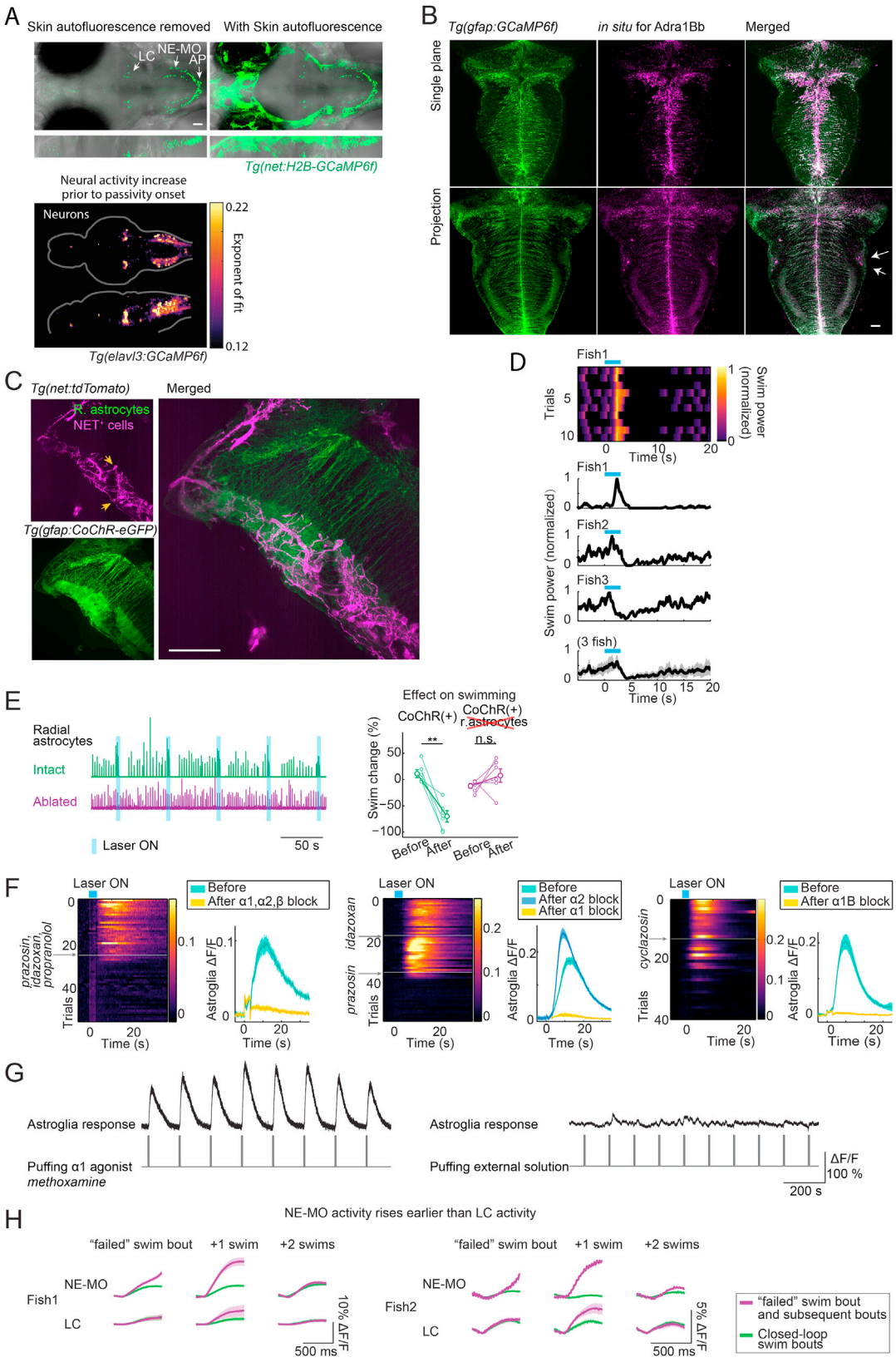
(C) Optogenetic stimulation during functional imaging of radial astrocytes shows that exciting these cells can cause spatially propagating calcium activity. Top row: after optogenetic excitation of astrocytic processes in the L-MO, a wave of calcium can be seen propagating from the site of excitation to the astrocytic cell bodies located along the midline. The time course of evoked activity resembles spontaneous activity of single astrocytic cells (top right), which suggests that optogenetic stimulation is sufficient to drive an endogenous calcium-dependent excitatory pathway in radial astrocytes. Scale bars, 20 μm.

(D and E) Optogenetic stimulation of L-MO astrocytic processes suppresses swimming. (D) *Top:* CoChR-negative controls showed no difference in behavior when the L-MO was stimulated or light was targeted outside of the brain (targeted to the ear, chosen because it is very close to the L-MO but outside the brain). A small response to the blue light was present if either the L-MO or the ear was targeted. *Bottom:* The same experiment in animals expressing CoChR in radial astrocytes showed a difference in behavior, with L-MO stimulation leading to a prolonged suppression of swimming. In total 30 pulses with a width of 10 ms and inter-pulse-interval of 90 ms were delivered for each stimulation (Methods). Shading represents SEM (E) Stimulation in Opto-α1-AR expressing animals similarly showed a suppression of swimming. A continuous 10 s of light stimulation was delivered for each stimulation (Methods). Shading represents SEM.

(F) GABAergic neurons in L-MO activated during switches to passivity. Left, fish switching to the passive state after entering open-loop, with activity increases in GABAergic neurons in L-MO at the time of switches. Right, summary of all GABAergic neurons in L-MO from 6 fish. Cells are ordered by the center of mass along the time axis. Shading represents SEM.

(G) Capsaicin activation of astrocytic calcium causes an increase in neuronal activity in L-MO neurons, consistent with the activation of GABAergic neurons in L-MO after astrocytic stimulation (Figure 5R). Shading represent SEM.

(H) Optogenetically stimulating radial astrocytes causes increases in neuronal activity in L-MO neurons. Shading represents SEM.



(legend on next page)

Figure S6. Calcium Signals in Noradrenergic Neurons Increase before Switches to Passivity, Noradrenergic Receptors Are Highly Expressed and Co-localized with Radial Astrocytes in the Lateral Hindbrain, Activation of Noradrenergic Neurons Can Elicit a Transient Increase in Motor Vigor, and Coding for Swim Failures in Different NE Nuclei, Related to Figures 6 and 7

(A) *Top:* The nuclei of noradrenergic (NE) neurons (Farrar et al., 2018) are labeled in *Tg(net:H2B-GCaMP6f)* fish. *Bottom:* The neurons which increase their activity fastest prior to passivity onset overlap with the regions where the somata of noradrenergic cells, including the locus coeruleus (LC) and the NE cluster of the medulla oblongata (NE-MO), are located. Imaged line in the bottom panel is *Tg(elavl3:GCaMP6f)*. For easier visualization of noradrenergic cells, skin auto-fluorescence was removed manually in the top-left panel.

(B) Adrenergic receptor 1Bb expression visualized through *in situ* hybridization. The lateral hindbrain (L-MO) showed high levels of adra1Bb receptor expression (arrows). The receptors can also be seen to line the astrocytic processes that run medial-lateral, consistent with RNA-seq results (see Table S2). The projection is over a 40 μm thick section.

(C) Maximum intensity projection over 50 μm at the level of the L-MO of a *Tg(gfap:CoChR-eGFP); Tg(net:tdTomato)* fish that was expanded 4× through expansion microscopy to show a dense network of noradrenergic fibers expressing the norepinephrine transporter (NET) intertwined with glial processes in the L-MO. The section also shows the cell bodies of noradrenergic neurons (yellow arrows) located at the along the length of the medulla showing that indeed, noradrenergic neurons of the medulla oblongata (NE-MO) project to the L-MO.

(D) Photostimulating NE-MO can briefly potentiate swimming. In closed-loop, optogenetically stimulating NE-MO first increased swim vigor before it suppressed swimming. *Top*, single trials from an example fish. *Bottom,* mean swim power change from 3 example fish and the average from these 3 fish. Shaded region represents SEM.

(E) Photostimulating NE-MO no longer suppresses swimming after ablation of radial astrocytes in L-MO.

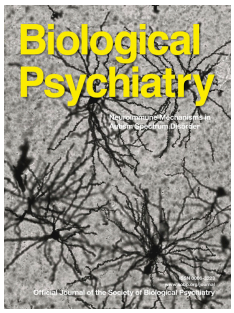
(F) Blocking α1 adrenergic receptors, but not α2 adrenergic receptors, abolishes the activation of radial astrocytes in L-MO by NE-MO stimulation. Optogenetic stimulation of NE-MO activates radial astrocytes in L-MO, and this activation was abolished by either adding a mixture of antagonists targeting multiple subtypes of adrenergic receptors (*left*, prazosin for α1, idazoxan for α2, propranolol for β), by an antagonist of α1 adrenergic receptors alone (*middle*), or by an antagonist of α1B adrenergic receptors alone (*right*), but not by an antagonist of α2 adrenergic receptors (*middle*). Shading represents SEM.

(G) Puffing methoxamine, an agonist of the α1 adrenergic receptor, triggered calcium responses in radial astrocytes in L-MO (*left*). Puffing external solution didn't trigger a response (*right*).

(H) Calcium activity in the NE-MO represents swim failures (Figure 7A). Calcium activity in the LC also represents swim failures, but only on the subsequent swim bout; this response may be downstream of the response in the NE-MO. The signals were aligned by subtracting the fluorescence right before the swim onset. Shading represents SEM.

Scale bar in (A-C): 50 μm.

Journal Pre-proof



Glucocorticoids promote fear generalization by increasing the size of a dentate gyrus engram cell population

Sylvie L. Lesuis, Niek Brosens, Nathalie Immerzeel, Rolinka J. van der Loo, Miodrag Mitrić, Pascal Bielefeld, Carlos P. Fitzsimons, Paul J. Lucassen, Steven A. Kushner, Michel C. van den Oever, Harm J. Krugers

PII: S0006-3223(21)01250-6

DOI: https://doi.org/10.1016/j.biopsych.2021.04.010

Reference: BPS 14534

To appear in: *Biological Psychiatry*

Received Date: 7 October 2020

Revised Date: 12 April 2021

Accepted Date: 13 April 2021

Please cite this article as: Lesuis S.L., Brosens N., Immerzeel N., van der Loo R.J., Mitrić M., Bielefeld P., Fitzsimons C.P., Lucassen P.J., Kushner S.A., van den Oever M.C. & Krugers H.J., Glucocorticoids promote fear generalization by increasing the size of a dentate gyrus engram cell population, *Biological Psychiatry* (2021), doi: https://doi.org/10.1016/j.biopsych.2021.04.010.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier Inc on behalf of Society of Biological Psychiatry.

Journal Pre-proof

Glucocorticoids promote fear generalization by increasing the size of a dentate gyrus engram cell population

Short title: Stress increases fear generalization and engram size

Authors

Sylvie L. Lesuis[1,2,†*], Niek Brosens[1,†], Nathalie Immerzeel[1], Rolinka J. van der Loo[3], Miodrag Mitrić[3], Pascal Bielefeld[1], Carlos P. Fitzsimons[1], Paul J. Lucassen[1], Steven A. Kushner[4], Michel C. van den Oever[3,‡], Harm J. Krugers[1,‡*]

Affiliations

[1] Brain Plasticity Group, SILS-CNS, University of Amsterdam, Amsterdam, The Netherlands

[2] Program in Neurosciences and Mental Health, Hospital for Sick Children, Toronto, Canada.

[3] Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research (CNCR), Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[4] Department of Psychiatry, Erasmus University Medical Center, Rotterdam, The Netherlands

***Corresponding authors:** S.L. Lesuis (Sylvie.lesuis@sickkids.ca) and H.J. Krugers (h.krugers@uva.nl) Brain Plasticity Group, SILS-CNS, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands

[†] These authors contributed equally: Sylvie L. Lesuis, Niek Brosens

[‡] Shared senior authors: Michel C. van den Oever, Harm J. Krugers

Key words (6): stress, glucocorticoids, engram, fear learning, memory generalization, excitability

Journal Pre-proof

Abstract

Background: Traumatic experiences, such as conditioned threat, are coded as enduring memories that are frequently subject to generalization, which is characterized by (re-) expression of fear in safe environments. However, the neurobiological mechanisms underlying threat generalization following a traumatic experience and the role of stress hormones in this process remain poorly understood.

Methods: We examined the influence of glucocorticoid hormones on the strength and specificity of conditioned fear memory at the level of sparsely distributed dentate gyrus (DG) engram cells in male mice.

Results: We report that elevating glucocorticoid hormones after fear conditioning induces a generalized contextual fear response. This was accompanied by a selective and persistent increase in the excitability and number of activated DG granule cells. Selective chemogenetic suppression of these sparse cells in the DG prevented glucocorticoid-induced fear generalization and restored contextual memory specificity, while leaving expression of auditory fear memory unaffected.

Conclusions: These results implicate the sparse ensemble of DG engram cells as a critical cellular substrate underlying fear generalization induced by glucocorticoid stress hormones.

Journal Pre-proof

Journal Pre-proof

Introduction

Stressful and emotionally arousing events are generally remembered well (*1*). Robust memory retention is highly adaptive but comes at increasing cost when memories lose their specificity. Memory encoding and/or processing under stressful circumstances can result in generalization of fear expression in which memories are expressed promiscuously, even in safe situations and/or in the absence of predictive cues, as exemplified by posttraumatic stress disorder (PTSD) or generalized anxiety disorders (*1*–*3*).

Stressful experiences activate the Hypothalamus-Pituitary-Adrenal (HPA)-axis and the release of glucocorticoid hormones (cortisol in humans, corticosterone in rodents) from the adrenal cortex (*4*). Glucocorticoid hormones alter neuronal function by activating high affinity brain mineralocorticoid receptors (MRs) and lower affinity glucocorticoid receptors (GRs) (*5*, *6*). Via these receptors, glucocorticoid hormones rapidly and persistently increase glutamatergic synaptic transmission, which is critical for synaptic plasticity and learning and memory (*6*–*9*). Behaviorally, glucocorticoid hormones facilitate memory consolidation, extinction and habitual learning (*1*, *10*, *11*). In addition, glucocorticoid hormones have also been reported to modify memory contextualization which may confer memory generalization (*12*, *13*).

Specific memories are encoded by small populations of cells, also known as ‘engram cells’, that are activated during specific learning epochs (*14*–*16*). The hippocampal dentate gyrus (DG) is strongly involved in spatial aspects of memory processing and learning-activated engram populations in this region support the consolidation and expression of contextual fear memories (*17*–*19*). Here, we examined whether glucocorticoid levels affect the specificity of contextual fear memory through changes in the size and physiological properties of the hippocampal DG engram cell population (*20*).

Materials and Methods

Study design

Wildtype (C57BL/6J; WT) mice (Harlan, The Netherlands) and Arc::dVenus mice that were backcrossed for more than 10 generations into C57BL/6J mice (*21*) were used. Experiments were performed during the light phase, using male adult mice (8–12 weeks). Mice were individually housed for 14 days prior to the experiments. All experiments were conducted under EU directive 2010/63/EU for animal experiments and were approved by the animal welfare committee of the University of Amsterdam. Mice were maintained under standard housing conditions (temperature 20-22 °C, 40-60% humidity) on a 12 h light/dark cycle with standard chow and water available *ad libitum*. Assignment to treatment condition, data collection and data processing was conducted at random, by an experimenter blind to the experimental condition. Sample size was determined *a priori* by conducting a power analysis. All experimental procedures were performed using independent cohorts of mice.

Fear conditioning

Fear training involved placing mice into a conditioning chamber, and, 3 min later, presenting three co-terminating presentations of a tone conditioned stimulus (CS) (30s, 20.8 kHz, 82 dB) and a foot shock unconditioned stimulus (US) (2s, 0.2 mA). This mild conditioning paradigm was selected to avoid potential ceiling effects on behavior after corticosterone injections (*22*, *23*). During testing mice were place in the same (context A) and/or a novel context (context B) and freezing was assessed. See Supplementary Materials and Methods for details.

Drug treatment

Corticosterone (Sigma) (dissolved in 5% EtOH, final dose: 2 mg/kg, injection volume: 5 µl/g body weight) was injected intraperitoneally (i.p.) immediately following fear conditioning. RU486 (Sigma) (dissolved in 10% EtOH in peanut oil, final dose: 10 mg/kg, injection volume:

5µl/g body weight) was injected i.p. immediately following conditioning. The corresponding vehicle solution was administered to control group accordingly (*22*).

Immunohistochemistry

Sections were stained for Arc and c-fos as described in the Supplementary Materials and Methods.

Recording Action Potential (AP) Frequency

Mice were left undisturbed for 1h after fear conditioning and injections. Mice were sacrificed by acute decapitation and brains were quickly removed and prepared for whole cell recordings in the DG in current clamp mode as described in the Suppl. M&M.

Recording miniature excitatory postsynaptic currents (mEPSCs)

Mice were left undisturbed for 5h after fear conditioning and injections. Subsequently, mice were sacrificed by acute decapitation and brains were quickly removed and prepared for mEPSC recordings in the DG as described in the Suppl. M&M.

Stereotactic microinjections for viral-TRAP

AAV-Fos::CreER^{T2} (titer: 1.2 x 10^{12}) and Cre-dependent AAV-hSyn:DIO-hM4Di-mCherry and AAV-hSyn::DIO-mCherry (titer: 5-6 x 10^{12}) were packaged as serotype 5 virus. A virus mixture of AAV5-Fos::CreER^{T2} and Cre-dependent AAV (ratio 1:500, AAV-Fos-CreER^{T2} was injected at a final titer of 2.4 x 10^{9}) (*24*, *25*) was infused bilaterally in the DG using microinjection glass needles (0.5 µl; flow rate: 0.1 µl/min; AP = -2.2; ML= ±1.2; DV= -2.2, relative to Bregma).

4-Hydroxytamoxifen (4OHT) treatment

4OHT (H6278, Sigma-Aldrich Chemie N.V., The Netherlands) (50 mg/ml 4OHT dissolved in DMSO (D8148, Sigma-Aldrich Chemie N.V., The Netherlands) and diluted 10x in saline containing 2% Tween80 (P1754, Sigma-Aldrich Chemie N.V., The Netherlands) and 10x in

saline; final concentration: 2.5 mg/ml 4OHT, 5% DMSO and 1% Tween80 in saline) was injected 2h after training in hM4Di-mCherry and control mice. The final dose was 25 mg/kg.

Chemogenetic intervention

Clozapine N-oxide (CNO, BML-NS105, Enzo LifeSciences, Brussels) (dissolved in saline; final dose: 5 mg/kg, injection volume 10 µl/g body weight) was injected i.p. 30 min before a retrieval session in both mCherry and hM4Di-mCherry mice.

Statistical analysis

Data were analyzed using SPSS 22.0 (IBM software). All data are expressed as mean + SEM. Data were considered statistically significant when $p<0.05$ and were tested two-sided. Outliers were determined using a Grubb's test, removing a maximum of 1 data point. Independent-samples t-tests were performed to assess differences between saline and corticosterone-treated mice. When assumption of normality was not met (based on Shapiro-Wilk test), Mann-Whitney test was conducted. A 2x2 ANOVA was performed to assess the interaction between two factors (treatment x context or treatment x cell type), with post-hoc Tukey testing.

Journal Pre-proof

Results

Corticosterone induced fear memory generalization

Mice underwent fear conditioning by 3 mild footshocks (0.2 mA) paired with a discrete tone, immediately followed by systemic administration of corticosterone (2 mg/kg) or saline (**Fig. 1A**). As expected, we observed a conditioning-induced increase of plasma corticosterone levels, which was even further elevated and prolonged in mice receiving post-training corticosterone administration (**Suppl. Fig. 1**).

Freezing was comparable between saline and corticosterone-injected mice during re-exposure to the training context 24 hours after conditioning (**Fig. 1B**) ($t(14)=0.24$, $p=0.82$). In contrast, mice receiving corticosterone immediately following conditioning exhibited

significantly higher freezing to a distinct context (neutral context B), in both the presence and absence of the conditioned tone (**Fig. 1C**) ($F(1,28)=77.13$, $p<0.0001$; post hoc: saline-baseline vs. CORT-baseline: $p=0.02$; saline-tone vs. CORT-tone: $p=0.002$). Moreover, corticosterone injection significantly increased contextual fear generalization, which was quantified as the ratio of context-evoked freezing in context B (novel context) relative to context A (training context) (**Fig. 1D**) ($t(14)=2.84$, $p=0.013$). Although absolute freezing to the tone was higher in corticosterone-treated mice, the relative increase in freezing during the tone compared to baseline freezing in context B was similar between saline and corticosterone-treated mice (**Fig. 1E**) ($t(14)=1.61$, $p=0.13$), indicating that the increase during the tone was caused by a generalized fear response in the novel context.

We next examined whether the effect of corticosterone on generalization of conditioned fear was mediated by activation of glucocorticoid receptors (GRs). Co-administration of the GR antagonist RU486 prevented the effect of corticosterone on freezing in a neutral context (**Fig. 1F**) ($F(1,28)=17.46$, $p=0.0003$, post hoc: Veh-Saline vs. veh-CORT: $p<0.0001$; Veh-CORT vs. RU486-CORT: $p<0.0001$) as well as to the tone (**Fig. 1G**) ($F(1,28)=5.89$, $p=0.02$, post hoc: Veh-Saline vs. veh-CORT: $p=0.004$; Veh-CORT vs. RU486-CORT: $p<0.0001$). Furthermore, freezing in context A and the enhanced freezing in neutral context B following corticosterone treatment was absent in mice that underwent the same paradigm, but without receiving foot shocks during exposure to context A (**Fig. 1H-K**). Together, these results show that increased corticosterone levels immediately after fear conditioning subsequently results in a generalized fear response in a neutral context. Moreover, we show that the effect of corticosterone is mediated through activation of GRs.

Corticosterone enhanced the size of a fear engram cell population

In order to identify how corticosterone triggered memory generalization, we examined neuronal activation in the DG, a brain region that is strongly involved in contextual fear memory (17–19). For this, mice underwent fear conditioning followed by corticosterone or

saline injection. Ninety minutes after conditioning, brains were isolated and immunostained for c-fos, an immediate early gene and molecular marker of recent neuronal activation (*26*) (**Fig. 2A**). The number of neurons expressing endogenous c-fos was increased in the DG of corticosterone-treated mice at 90 minutes after training compared to saline-treated animals (**Fig. 2B,C**) ($t(10)=7.8$, $p<0.0001$). To further characterize the neuronal population that was activated by fear conditioning, we used a transgenic mouse with destabilized Venus fluorescent protein (dVenus) expression under control of the Activity Regulated Cytoskeleton Associated Protein (Arc) promotor (Arc::dVenus mice) (21), which allows for visualization and characterization of activated neurons in the DG for at least 24 hours after conditioning (17). At 90 minutes after fear conditioning, we observed an increase in the number of dVenus labeled cells in corticosterone-treated animals (**Fig. 2D,E**) ($t(14)=3.0$, $p=0.009$). After context exposure alone (without the delivery of a foot shock), corticosterone did not increase the number of c-fos or dVenus labeled cells in the DG compared to control mice (c-fos: $t(14)=1.8$, $p=0.10$; dVenus: $t(14)=0.06$, $p=0.95$) (**Fig. 2F-J**). Ninety minutes after fear conditioning, we observed significant colocalization of cells expressing c-fos and endogenous Arc ($t(15)=22.55$, $p<0.0001$), as well as of the cells co-expressing c-fos and dVenus ($t(14)=18.6$, $p<0.0001$) (**Fig. 2K-M**), indicating that these molecular markers represent, to a large extent, the same cellular population, consistent with what has been reported previously (17). Interestingly, we found that the number of Arc::dVenus positive cells was still increased in corticosterone-treated mice at 24 hours after training compared to control animals (**Fig. 2N-P**) ($t(30)=4.92$, $p<0.001$), suggesting that elevated corticosterone levels during the first hours after conditioning persistently increased the size of the DG engram population.

Upon retrieval (**Fig. 2Q**), corticosterone-injected and saline-injected mice showed comparable numbers of c-fos$^+$ cells in context A (**Fig. 2R**) ($F(1,25)=10.58$, $p=0.003$; post hoc: saline-Context A vs. CORT-Context A: $p=0.99$). In contrast, corticosterone-injected mice exhibited a larger number of c-fos$^+$ cells than saline-injected mice in context B (**Fig. 2R**)

(saline-Context B vs. CORT-Context B: p<0.001). To assess whether training-activated neurons were re-activated upon exposure to context A or B, we used the long-lasting detectability of dVenus in the DG to examine re-activation during memory retrieval. Importantly, at 90 mins after memory retrieval in either context A or context B (**Fig. 2Q**), the density of dVenus$^+$ neurons was equivalent to the density in the absence of a retrieval session (see **Fig. 2P**; main treatment effect: F(1,40)=26.73, p=<0.0001; treatment x test interaction effect: F(2,40)=0.17, p=0.84) (**Fig. 2S**), indicating that retrieval did not alter the size of the training-induced dVenus$^+$ population. In context A and B, we observed preferential expression of c-fos within the dVenus$^+$ populations compared with the dVenus$^-$ populations after saline and corticosterone treatment (main effect cell type: context A: F(1,24)=578.8, p<0.0001; context B: F(1,24)=320.5, p<0.0001). However, corticosterone-injected mice exhibited increased colocalization of c-fos$^+$ and dVenus$^+$ neurons in the DG after exposure to context B compared with saline-treated mice (treatment x cell type interaction effect: F(1,24)=14.22, p<0.001, post hoc: saline-dVenus$^+$ vs. CORT-dVenus$^+$: p<0.0001) (**Fig. 2T,V**), whereas this did not differ between groups after exposure to context A (treatment x cell type interaction effect: F(1,24)=1.03, p=0.32) (**Fig. 2T-U**). Importantly, colocalization between c-fos$^+$ and dVenus$^+$ cells was significantly higher than chance level in Context A for both groups, and higher than chance level in Context B for the CORT-treated animals (**Suppl. Fig. 2**). These findings are consistent with a less precise contextual representation of the memory engram and fear generalization resulting from post-training corticosterone.

Corticosterone selectively enhanced the excitability of dVenus$^+$ neurons after fear conditioning

Since neural activation and excitability determine neuronal selection into a memory engram (*17*, *27–29*), we used whole-cell patch clamp recordings to further investigate the effects of corticosterone on the physiological properties of activated dVenus$^+$ neurons and non-activated (dVenus$^-$) neighboring neurons at 1 h after conditioning. Corticosterone injection

after training did not affect membrane resistance, input resistance, minimum current intensity, membrane capacitance or access resistance differently in dVenus$^-$ and dVenus$^+$ neurons (**Suppl. Fig. 3**). Depolarizing current injections evoked an increase in action potential (AP) frequency, which was significantly reduced in dVenus$^+$ neurons compared to dVenus$^-$ neurons (**Fig. 3A,B**) (current injected x treatment x cell type interaction effect: $F(75,1225)=1.93$, $p<0.0001$; saline-dVenus$^+$ vs. saline-dVenus$^-$: $p=0.04$). However, compared with saline-injected animals, corticosterone selectively increased the AP frequency elicited by depolarizing current injections in dVenus$^+$, but not in dVenus$^-$, neurons in mice that were fear conditioned (**Fig. 3A-C**) (Saline-dVenus$^+$ vs. CORT-dVenus$^+$: $p=0.26$; CORT-dVenus$^+$ vs. CORT-dVenus$^-$: $p=0.78$). In naïve home cage mice with no prior fear conditioning, the AP frequency was similarly lower in dVenus$^+$ cells compared with dVenus$^-$ cells in both groups, but corticosterone injection did not enhance AP frequency in dVenus$^+$ neurons (nor in dVenus$^-$ neurons) (**Fig. 3D,E**) (current injected x treatment x cell type interaction effect: $F(75,1700)=4.30$, $p<0.0001$; saline-dVenus$^+$ vs. saline-dVenus$^-$: $p=0.01$; saline-dVenus$^+$ vs. CORT-dVenus$^+$: $p=0.93$). Hence, corticosterone increased the excitability of activated (dVenus$^+$) DG neurons, potentially contributing to the enhanced ensemble size at 90 min and 24 h after conditioning (see **Fig. 2B-E**).

At 5 hours after training, we examined the effects of corticosterone injection on synaptic transmission by analyzing the properties of miniature excitatory postsynaptic currents (mEPSCs). We found that the frequency of mEPSCs was higher in dVenus$^+$ neurons compared to dVenus$^-$ neurons of saline injected mice (**Fig. 3F-H**) ($F(1,14)=10.9$, $p=0.005$). Compared with saline-treated mice, corticosterone injection significantly enhanced the mEPSC frequency selectively in dVenus$^-$ neurons (**Fig. 3F-H**) (interaction effect: $F(1,14)=8.05$, $p=0.01$, post-hoc: $p=0.04$), but did not alter the amplitude of mEPSCs in dVenus$^+$ and dVenus$^-$ neurons (**Fig. 3G,H**) ($F(1,14)=2.91$, $p=0.11$). The increase in frequency of mEPSCs in dVenus$^-$ neurons after corticosterone injection may reflect altered presynaptic input onto these neurons.

Chemogenetic suppression of the DG engram population reduced corticosterone-induced generalized fear

Previous studies have shown that inhibition of fear conditioning-activated DG neurons impairs memory retrieval upon re-exposure to the training context (*27*, *31*). Based on this, we next investigated whether the DG neurons activated during conditioning paired with corticosterone were responsible for the observed fear generalization in a neutral context. To assess this, we used a well-established viral-TRAP (Targeted Recombination in Activated Populations) approach, restricting expression of the inhibitory Gi-DREADD (hM4Di (*30*)) to activated neurons under control of the Fos promoter in a Cre recombinase and 4-hydroxytamoxifen (4OHT) dependent manner (*24*, *25*) (**Fig. 4A**). After training, both groups were injected with corticosterone and 2 h later received 4OHT to allow expression of hM4Di-mCherry or mCherry alone (control) in activated DG neurons (**Fig. 4B**). When mice were re-exposed to context A (without CNO) freezing levels were similar between groups **(Fig. 4C**) ($t(15)=0.64$, $p=0.53$). Thirty minutes before exposure to the neutral context B, all mice received CNO to suppress the activity of DG neurons that were tagged after training. Whereas mCherry control mice showed substantial baseline freezing in context B, suppression of hM4Di$^+$ neurons reduced freezing behavior (**Fig. 4D**) (post hoc: baseline-mCherry vs. baseline-hM4d: $p=0.0001$) to a level that that was comparable to control mice that did not receive post-training corticosterone (see **Fig. 1B**). Indeed, contextual fear generalization, quantified as the ratio of context-evoked freezing in context B (novel context) relative to freezing in context A (training context), was reduced following suppression of hM4Di$^+$ neurons in corticosterone treated mice (**Fig. 4E**) ($t(15)=3.55$, $p=0.003$). Freezing to the tone remained unaffected by CNO (**Fig. 4D**) ($F(1,30)=7.57$, $p=0.01$, post hoc tone-mCherry vs. tone-hM4Di: $p=0.66$). This confirms that inhibition of these DG neurons prevented the expression of contextual memory generalization.

Finally, we determined whether DG neurons that were tagged with viral-TRAP were reactivated upon exposure to context B, similar to what we observed in the Arc::dVenus

mice. Indeed, control mice expressing mCherry alone showed substantial reactivation of tagged neurons, whereas in mice injected with hM4Di-mCherry, c-fos expression was selectively reduced in mCherry^{+} neurons (**Fig. 4F,G**) ($F(1,30)=48.79$, $p<0.0001$, post hoc: control-mCherry^{+} vs. hM4Di-mCherry^{+}: $p<0.0001$; hM4Di-mCherry^{-} vs hM4Di-mCherry^{+}: $p=0.86$). This confirms that the neurons tagged during training were indeed reactivated during expression of fear in a neutral context. Together, these data indicate that corticosterone-induced generalized fear is driven by the increased DG engram population and that suppression of this improves contextual discrimination.

Journal Pre-proof

Journal Pre-proof

Discussion

A delicate balance between memory strength, specificity and generalization is essential to enable efficient encoding and retrieval of highly salient, stress-related information (*32*). We report here that corticosterone, the main glucocorticoid stress hormone in mice, alters the size of a fear memory-encoding DG engram population and that this population is functionally responsible for the expression of generalized contextual fear.

Various lines of evidence indicate that GRs are involved in enhancing memory consolidation. Our findings demonstrate that glucocorticoid hormones, via GRs, are also involved in the effects of corticosterone on memory generalization (**Fig. 1F,G**). In line with this, it was recently reported that corticosterone, at levels that activate GRs, reduces memory accuracy (*33*). Thus, activation of GRs not only enhances contextual memory consolidation, but also memory generalization, and we now show that this effect is mediated via an increase in the DG engram cell population. The presently observed contextual memory generalization in a novel context was not simply due to enhanced memory strength, since no differences were observed between saline and corticosterone injected mice in fear expression when mice were tested in the conditioning context. Additionally, the observed effects required associative learning, since corticosterone did not alter the number and excitability of activated DG neurons when mice did not receive aversive foot-shocks in the conditioning context. These two findings point toward a specific role of corticosterone in the enhancement of memory generalization.

We observed that corticosterone increased the sparse number of activated DG cells after training, as well as at 24 h after training, pointing to a stable change in the fear-encoding DG engram population. Whereas reactivation of learning-activated DG neurons upon re-exposure to the conditioned context was not affected by corticosterone treatment, we found enhanced reactivation of the same population upon exposure to a neutral context. Importantly, selective inhibition of DG neurons that are activated by fear conditioning plus elevated corticosterone levels prevented the expression of generalized contextual fear, while

leaving memory for the conditioned tone unaffected. Although we did not examine whether activity of the learning-activated DG neurons was required for memory expression upon re-exposure to the conditioning context (i.e. context A), this has previously been shown by others (27, 31) and we thus expect that this is not further affected by elevated corticosterone levels. However, we extend these earlier findings by showing here that the corticosterone-induced increase in the size of a DG engram cell population is responsible for a generalized expression of fear also in a neutral context. Our results support previous observations that memory specificity is related to the size of neuronal ensembles in the DG (*34*–*36*), but for the first time we causally demonstrate that glucocorticoid hormones have a critical role in modulating the incorporation of neurons in an engram population.

Alterations in neuronal excitability have been implicated in recruitment of neurons in a memory ensemble in the lateral amygdala (*29*, *37*) and after retrieval in the DG (*27*), possibly via regulation of potassium channels (*17*, *38*). We found that fear conditioning-activated DG neurons have lower excitability compared to non-activated neurons, independent of elevated glucocorticoid hormone levels. However, we demonstrate that glucocorticoid hormones specifically enhance excitability of dVenus$^+$ neurons after fear conditioning, while leaving excitability unaffected in the absence of aversive associative learning. The reduced excitability of the dVenus$^+$ cells that we observed in control mice after learning contrasts with enhanced excitability of dVenus$^+$ cells in the LA after fear conditioning (*37*) and with a transient increase in excitability of DG engram cells after memory retrieval (*38*). We speculate that decreased dVenus$^+$ cell excitability shortly after learning functions to protect the integrity of the information encoded in the DG engram cell population by avoiding subsequent recruitment of these neurons during a different experience shortly thereafter and thereby the linking of memories (*39*, *40*). This would be consistent with the pattern separation function of the DG (*41*). At the cellular level, the reduced excitability is likely due to a decrease in the membrane resistance in dVenus$^+$ cells (**Suppl. Fig. 3A**). What the molecular mechanism is that underlies the change in membrane

resistance is an important topic for future research, but is likely driven by post-translational modifications or trafficking of membrane channels given the rapid induction after learning (within 1 h). Changes in potassium channels may contribute to conditioning-induced reduction in excitability (*42*), because these proteins are thought to modulate the excitability of fear memory-encoding DG granule cells (*17*, *43*). As corticosterone selectively enhanced excitability of dVenus$^+$ neurons, glucocorticoid hormones may act as a switch to enhance excitability of conditioning-activated DG neurons and thereby the stable incorporation of neurons in a fear memory-encoding engram cell population.

Persistent alterations in synaptic strength in engram cells have been implicated in memory consolidation and this is crucial for subsequent memory expression (*37*, *44*). In control animals, we observed that fear conditioning elicited an increase in the mEPSC frequency specifically in activated DG neurons. This is in line with enhanced input of projections from the medial entorhinal cortex onto Fos-expressing DG neurons after fear conditioning (*45*). Although our recordings were performed on neurons that expressed dVenus driven by the Arc promoter, we found that the Arc$^+$ and Fos$^+$ population overlap to a large extend (*17*). Moreover, Fos-expressing DG neurons drive generalized fear in a neutral context (*45*), similar to what we observed with our Fos-promoter based viral-TRAP approach. We also found that glucocorticoid hormones enhanced the frequency of mEPSCs in non-activated DG neurons to the level detected in activated neurons 5 h after training. Thus, glucocorticoids seem to increase general synaptic transmission in DG neurons after training, possibly via increased input from the medial entorhinal cortex (*45*), which may contribute to aberrant processing of contextual information (*46*).

In line with our observations, it has been reported that stress elicits memory generalization and that glucocorticoid hormones impair the ability to correctly predict threat (*3*, *12*, *13*). Although some level of memory generalization is adaptive, as it allows a learned response to transfer to other relevant situations and stimuli, overgeneralization of fear is often maladaptive. Indeed, stressful learning experiences contribute to fear sensitization, resulting

in exaggerated startle responses to trauma-unrelated stimuli (*47*), particularly in relation to the contextual component of the memory trace. This impairment is similar to symptoms observed in PTSD patients, who often forget contextual peri-traumatic cues, whereas salient but irrelevant stimuli are strongly remembered. These salient cues, and other similar cues, can then induce strong fear responses in contexts that differ from the traumatic environment (*2*). Interestingly, noradrenaline, another key mediator of the stress response, has been shown to contribute to memory specificity (*48*). Potentially, a closely governed balance in the release of noradrenaline (specificity) and glucocorticoids (generalization), which have been reported to interact at the level of the neuronal structure (*49*), electrophysiological functioning (*10*, *50*) and behavior (*51*), is essential for accurate memory consolidation and retrieval.

The present study demonstrates that a transient post-learning increase in glucocorticoid hormone levels in male mice reduces memory specificity, thereby promoting fear memory generalization. We further reveal that glucocorticoids alter the size of the DG neuronal population that is activated by an aversive experience. Selectively suppressing this neuronal ensemble restores memory specificity. Given that stress-related disorders are often sex-dependent and that glucocorticoids regulate fear memory formation in a sex-dependent manner (22), further investigation is warranted regarding the role of these hormones on memory generalization in females.

Moreover, understanding how glucocorticoids modify the delicate balance between memory strength and specificity through changes in the size of the engram cell population opens novel avenues for the development of treatments for stress-related disorders that are characterized by maladaptive aversive memories. In this respect, our current findings may be particularly relevant in relation to anxiety disorder and post-traumatic stress disorder which have been linked to memory generalization and impaired pattern separation (*52*).

Acknowledgments: The authors want to thank dr. Priyanka Rao-Ruiz (CNCR, VU) for helpful discussions during the design of experiments.

Disclosures: The authors report no biomedical financial interests or potential conflicts of interest.

Author contribution: SLL, MCvdO, and HJK contributed to conception and design of the study; SLL organized the database; SLL, RJvdL, PB, NB, MM and NI conducted the experiments, SLL performed the statistical analysis; SLL wrote the first draft of the manuscript; SLL, MCvdO, PB, CPF, NI, MM, PJL and HJK contributed to manuscript revision, read and approved the submitted version.

Data availability: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Journal Pre-proof

References

1. D. de Quervain, L. Schwabe, B. Roozendaal, Stress, glucocorticoids and memory: implications for treating fear-related disorders. *Nat. Rev. Neurosci.* **18**, 7–19 (2017).

2. APA, *Diagnostic and Statistical Manual of Mental Disorders IV—Text Revision* (American Psychiatric Press, Washington, DC, 2000).

3. J. E. Dunsmoor, A. R. Otto, E. A. Phelps, Stress promotes generalization of older but not recent threat memories. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9218–9223 (2017).

4. M. Joëls, T. Z. Baram, The neuro-symphony of stress. *Nat. Rev. Neurosci.* **10**, 459–466 (2009).

5. M. Joëls, R. A. Sarabdjitsingh, H. Karst, Unraveling the time domains of corticosteroid hormone influences on brain activity: rapid, slow, and chronic modes. *Pharmacol. Rev.* **64**, 901–38 (2012).

6. H. J. Krugers, C. C. Hoogenraad, L. Groc, Stress hormones and AMPA receptor trafficking in synaptic plasticity and memory. *Nat. Rev. Neurosci.* **11**, 675–681 (2010).

7. L. Groc, D. Choquet, F. Chaouloff, The stress hormone corticosterone conditions AMPAR surface trafficking and synaptic potentiation. *Nat Neurosci*. **11**, 868–870 (2008).

8. H. Karst, S. Berger, M. Turiault, F. Tronche, G. Schütz, M. Joëls, Mineralocorticoid receptors are indispensable for nongenomic modulation of hippocampal glutamate transmission by corticosterone. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 19204–7 (2005).

9. E. Y. Yuen, W. Liu, I. N. Karatsoreos, Y. Ren, J. Feng, B. S. McEwen, Z. Yan, Mechanisms for acute stress-induced enhancement of glutamatergic transmission and working memory. *Mol. Psychiatry*. **16**, 156–170 (2011).

10. M. Joëls, Z. Pu, O. Wiegert, M. S. Oitzl, H. J. Krugers, Learning under stress: how

does it work? *Trends Cogn. Sci.* **10**, 152–158 (2006).

11. L. Schwabe, O. T. Wolf, M. S. Oitzl, Memory formation under stress: Quantity and quality. *Neurosci. Biobehav. Rev.* **34** (2010), pp. 584–591.

12. N. Kaouane, Y. Porte, M. Vallée, L. Brayda-Bruno, N. Mons, L. Calandreau, A. Marighetto, P. V. Piazza, A. Desmedt, Glucocorticoids can induce PTSD-like memory impairments in mice. *Science*. **335**, 1510–3 (2012).

13. A. Shaam Al Abed, E. Ducourneau, C. Bouarab, A. Sellami, A. Marighetto, A. Desmedt, Preventing and treating PTSD-like memory by trauma contextualization. *Nat. Commun.* **11**, 4220 (2020).

14. S. A. Josselyn, S. Tonegawa, Memory engrams: Recalling the past and imagining the future. *Science (80-.).* **367** (2020), , doi:10.1126/science.aaw4325.

15. S. Tonegawa, M. D. Morrissey, T. Kitamura, The role of engram cells in the systems consolidation of memory. *Nat. Rev. Neurosci.* **19**, 485–498 (2018).

16. T. Kitamura, S. K. Ogawa, D. S. Roy, T. Okuyama, M. D. Morrissey, L. M. Smith, R. L. Redondo, S. Tonegawa, Engrams and circuits crucial for systems consolidation of a memory. *Science (80-.).* **356**, 73–78 (2017).

17. P. Rao-Ruiz, J. J. Couey, I. M. Marcelo, C. G. Bouwkamp, D. E. Slump, M. R. Matos, R. J. van der Loo, G. J. Martins, M. van den Hout, W. F. van IJcken, R. M. Costa, M. C. van den Oever, S. A. Kushner, Engram-specific transcriptome profiling of contextual memory consolidation. *Nat. Commun.* **10**, 1–14 (2019).

18. C. A. Denny, M. A. Kheirbek, E. L. Alba, K. F. Tanaka, R. A. Brachman, K. B. Laughman, N. K. Tomm, G. F. Turi, A. Losonczy, R. Hen, Hippocampal memory traces are differentially modulated by experience, time, and adult neurogenesis. *Neuron*. **83**, 189–201 (2014).

19. X. Liu, S. Ramirez, P. T. Pang, C. B. Puryear, A. Govindarajan, K. Deisseroth, S.

Tonegawa, Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*. **484**, 381–385 (2012).

20. T. Sasaki, V. C. Piatti, E. Hwaun, S. Ahmadi, J. E. Lisman, S. Leutgeb, J. K. Leutgeb, Dentate network activity is necessary for spatial working memory by supporting CA3 sharp-wave ripple generation and prospective firing of CA3 neurons. *Nat. Neurosci.* **21**, 258–269 (2018).

21. M. Eguchi, S. Yamaguchi, In vivo and in vitro visualization of gene expression dynamics over extensive areas of the brain. *Neuroimage*. **44**, 1274–1283 (2009).

22. S. Lesuis, L. A. E. Catsburg, P. J. Lucassen, H. J. Krugers, Effects of corticosterone on mild auditory fear conditioning and extinction; role of sex and training paradigm. *Learn. Mem.* **25**, 544–549 (2018).

23. C. Fitzsimons, L. van Hooijdonk, M. Schouten, I. Zalachoras, V. Brinks, T. Zheng, T. Schouten, D. Saaltink, T. Dijkmans, D. Steindler, J. Verhaagen, F. Verbeek, P. Lucassen, E. de Kloet, O. Meijer, H. Karst, M. Joels, M. Oitzl, E. Vreugdenhil, Knockdown of the glucocorticoid receptor alters functional integration of newborn neurons in the adult hippocampus and impairs fear-motivated behavior. *Mol. Psychiatry*. **18**, 993–1005 (2012).

24. E. Visser, M. R. Matos, R. J. van der Loo, N. J. Marchant, T. J. de Vries, A. B. Smit, M. C. van den Oever, A persistent alcohol cue memory trace drives relapse to alcohol seeking after prolonged abstinence. *Sci. Adv.* **6**, eaax7060 (2020).

25. M. R. Matos, E. Visser, I. Kramvis, R. J. van der Loo, T. Gebuis, R. Zalm, P. Rao-Ruiz, H. D. Mansvelder, A. B. Smit, M. C. van den Oever, Memory strength gates the involvement of a CREB-dependent cortical fear engram in remote memory. *Nat. Commun.* **10** (2019), doi:10.1038/s41467-019-10266-1.

26. F. C. Cruz, F. Javier Rubio, B. T. Hope, Using c-fos to study neuronal ensembles in

corticostriatal circuitry of addiction. *Brain Res.* **1628**, 157–173 (2015).

27. S. Park, E. E. Kramer, V. Mercaldo, A. J. Rashid, N. Insel, P. W. Frankland, S. A. Josselyn, Neuronal Allocation to a Hippocampal Engram. *Neuropsychopharmacology*. **41**, 2987–2993 (2016).

28. J.-H. Han, S. A. Kushner, A. P. Yiu, C. J. Cole, A. Matynia, R. A. Brown, R. L. Neve, J. F. Guzowski, A. J. Silva, S. A. Josselyn, Neuronal Competition and Selection During Memory Formation. *Science (80-.).* **316**, 457–460 (2007).

29. A. P. Yiu, V. Mercaldo, C. Yan, B. Richards, A. J. Rashid, H.-L. L. Hsiang, J. Pressey, V. Mahadevan, M. M. Tran, S. A. Kushner, M. A. Woodin, P. W. Frankland, S. A. Josselyn, Neurons are recruited to a memory trace based on relative neuronal excitability immediately before training. *Neuron*. **83**, 722–35 (2014).

30. S. Ferguson, P. Phillips, B. Roth, J. Wess, J. Neumaier, Direct-pathway striatal neurons regulate the retention of decision-making strategies. *J Neurosci*. **33**, 11668–11676 (2013).

31. C. A. Denny, M. A. Kheirbek, E. L. Alba, K. F. Tanaka, R. A. Brachman, K. B. Laughman, N. K. Tomm, G. F. Turi, A. Losonczy, R. Hen, Hippocampal memory traces are differentially modulated by experience, time, and adult neurogenesis. *Neuron*. **83**, 189–201 (2014).

32. A. Asok, E. R. Kandel, J. B. Rayman, The neurobiology of fear generalization. *Front. Behav. Neurosci.* **12** (2019), p. 329.

33. B. Roozendaal, G. Mirone, Opposite effects of noradrenergic and glucocorticoid activation on accuracy of an episodic-like memory. *Psychoneuroendocrinology*. **114** (2020), doi:10.1016/j.psyneuen.2020.104588.

34. T. Stefanelli, C. Bertollini, C. Lüscher, D. Muller, P. Mendez, Hippocampal somatostatin interneurons control the size of neuronal memory ensembles. *Neuron*.

89, 1074–1085 (2016).

35. N. Guo, M. E. Soden, C. Herber, M. T. Kim, A. Besnard, P. Lin, X. Ma, C. L. Cepko, L. S. Zweifel, A. Sahay, Dentate granule cell recruitment of feedforward inhibition governs engram maintenance and remote memory generalization. *Nat. Med.* **24**, 438–449 (2018).

36. J.-I. Kim, H.-Y. Cho, J.-H. Han, B.-K. Kaang, Which neurons will be the engram - activated neurons and/or more excitable neurons? *Exp. Neurobiol.* **25**, 55–63 (2016).

37. L. A. Gouty-Colomer, B. Hosseini, I. M. Marcelo, J. Schreiber, D. E. Slump, S. Yamaguchi, A. R. Houweling, D. Jaarsma, Y. Elgersma, S. A. Kushner, Arc expression identifies the lateral amygdala fear memory trace. *Mol. Psychiatry*. **21**, 364–375 (2016).

38. M. Pignatelli, T. J. Ryan, D. S. Roy, C. Lovett, L. M. Smith, S. Muralidhar, S. Tonegawa, Engram Cell Excitability State Determines the Efficacy of Memory Retrieval. *Neuron*. **101**, 274-284.e5 (2019).

39. D. J. Cai, D. Aharoni, T. Shuman, J. Shobe, J. Biane, W. Song, B. Wei, M. Veshkini, M. La-Vu, J. Lou, S. E. Flores, I. Kim, Y. Sano, M. Zhou, K. Baumgaertel, A. Lavi, M. Kamata, M. Tuszynski, M. Mayford, P. Golshani, A. J. Silva, A shared neural ensemble links distinct contextual memories encoded close in time. *Nature*. **534**, 115–118 (2016).

40. A. J. Rashid, C. Yan, V. Mercaldo, H.-L. L. Hsiang, S. Park, C. J. Cole, A. De Cristofaro, J. Yu, C. Ramakrishnan, S. Y. Lee, K. Deisseroth, P. W. Frankland, S. A. Josselyn, Competition between engrams influences fear memory formation and recall. *Science*. **353**, 383–7 (2016).

41. J. K. Leutgeb, S. Leutgeb, M. B. Moser, E. I. Moser, Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science (80-.).* **315**, 961–966 (2007).

42. N. A. Muma, S. G. Beck, Corticosteroids alter G protein inwardly rectifying potassium channels protein levels in hippocampal subfields. *Brain Res.* **839**, 331–335 (1999).

43. P. Rao-Ruiz, J. Yu, S. Kushner, S. Josselyn, Neuronal competition: microcircuit mechanisms define the sparsity of the engram. *Curr. Opin. Neurobiol.* **54**, 163–170 (2018).

44. T. J. Ryan, D. S. Roy, M. Pignatelli, A. Arons, S. Tonegawa, Engram cells retain memory under retrograde amnesia. *Science*. **348**, 1007–13 (2015).

45. X. Sun, M. J. Bernstein, M. Meng, S. Rao, A. T. Sørensen, L. Yao, X. Zhang, P. O. Anikeeva, Y. Lin, Functionally Distinct Neuronal Ensembles within the Memory Engram. *Cell*. **181**, 410-423.e17 (2020).

46. R.-M. Vouimba, G. Richter-Levin, Different patterns of amygdala priming differentially affect dentate gyrus plasticity and corticosterone, but not CA1 plasticity. *Front. Neural Circuits*. **7**, 80 (2013).

47. R. K. Pitman, A. M. Rasmusson, K. C. Koenen, L. M. Shin, S. P. Orr, M. W. Gilbertson, M. R. Milad, I. Liberzon, Biological studies of post-traumatic stress disorder. *Nat. Rev. Neurosci.* **13**, 769–787 (2012).

48. E. Atucha, V. Vukojevic, R. V. Fornari, G. Ronzoni, P. Demougin, F. Peter, P. Atsak, M. W. Coolen, A. Papassotiropoulos, J. L. McGaugh, D. J.-F. de Quervain, B. Roozendaal, Noradrenergic activation of the basolateral amygdala maintains hippocampus-dependent accuracy of remote memory. *Proc. Natl. Acad. Sci.* **114**, 9176–9181 (2017).

49. S. L. Lesuis, W. Timmermans, P. J. Lucassen, C. C. Hoogenraad, H. J. Krugers, Glucocorticoid and β-adrenergic regulation of hippocampal dendritic spines. *J. Neuroendocrinol.* **32** (2020), doi:10.1111/jne.12811.

50. M. Zhou, E. H. M. Bakker, E. H. Velzing, S. Berger, M. Oitzl, M. Joëls, H. J. Krugers,

Both mineralocorticoid and glucocorticoid receptors regulate emotional memory in mice. *Neurobiol. Learn. Mem.* **94**, 530–537 (2010).

51. B. Roozendaal, G. K. Hui, I. R. Hui, D. J. Berlau, J. L. McGaugh, N. M. Weinberger, Basolateral amygdala noradrenergic activity mediates corticosterone-induced enhancement of auditory fear conditioning. *Neurobiol. Learn. Mem.* **86**, 249–55 (2006).

52. M. A. Kheirbek, K. C. Klemenhagen, A. Sahay, R. Hen, Neurogenesis and generalization: A new approach to stratify and treat anxiety disorders. *Nat. Neurosci.* **15** (2012), pp. 1613–1620.

Journal Pre-proof

Figure legends

Figure 1. Corticosterone induced contextual fear memory generalization.

A. Schematic overview of the fear conditioning paradigm. Mice received corticosterone (CORT, 2 mg/kg) and/or RU486 (40 mg/kg) immediately after training. B**.** Following paired training, freezing in conditioning context A was comparable between experimental groups. N = 8 mice/group). **C.** Baseline freezing responses in the neutral context B and to the tone were enhanced after corticosterone treatment. N = 8 mice/group. **D.** Corticosterone-treated mice displayed more freezing in context B relative to context A. **E.** The ratio of freezing to the tone over baseline freezing in context B was not affected by corticosterone. **F.** Blocking the GR by RU486 treatment reduced corticosterone-induced freezing in context B. N = 8 mice/group. **G.** Blocking the GR by RU486 treatment reduced corticosterone-induced freezing to the tone. **H.** Response to context A of mice treated with saline or corticosterone after training in context A in the absence of foot-shocks. No difference in freezing levels was observed (N = 8 mice/group). **I.** Baseline freezing responses in the neutral context B and to the tone were not different between saline and corticosterone treated mice. N = 8 mice/group. **J.** No differences in the ratio of freezing between context A and B was present. **K.** No differences in the ratio between context B and tone freezing was present. Data are means + s.e.m. Statistical analysis was performed using a one-way repeated measures ANOVA with post-hoc Tukey test in **C** and **I**, a student's unpaired t-test in **B, D, E, H, J** and **K**, and a two-way ANOVA with post-hoc Tukey test in f and g. *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$.

Journal Pre-proof

Figure 2. Corticosterone enhanced the size of a fear conditioning-activated DG ensemble

A. Schematic overview of the fear conditioning paradigm. Mice were sacrificed 90 min after conditioning. CORT = corticosterone. **B.** Representative image showing c-fos expression in the DG. Scale bar=150 µm. **C.** CORT treatment increased the number of c-fos$^+$ cells in the DG after training. N = 6 mice/group. **D**. Representative image showing dVenus expression in de DG. Scale bar=250 µm. **E**. CORT treatment increased the number of dVenus$^+$ cells in the DG after training. N=8 mice/group. **F.** Schematic overview of the experiment. Mice were put in Context A without tone or foot shock, and were sacrificed 90 min after saline/CORT treatment. **G.** Representative image showing c-fos expression in the DG. Scale bar=150 µm**.** **H.** CORT treatment did not affect the number of c-fos$^+$ cells in the DG. N = 8 mice/group. **I.** Representative image showing dVenus expression in the DG. Scale bar=250 µm. **J.** CORT treatment did not affect the number of dVenus$^+$ cells in the DG. N = 8 mice/group. **K.** Representative images showing colocalization between c-fos$^+$ and arc$^+$ neurons, and between c-fos$^+$ and dVenus$^+$ neurons. Arrowhead indicate colocalization. Scale = 50 µm. **L.** Quantification of **K** (upper panels). N = 9 mice/group. **M.** Quantification of **K** (lower panels). N = 8. **N.** Schematic overview of the fear conditioning paradigm. Mice were sacrificed 24 h after conditioning. **O.** Representative images showing dVenus expression in saline and corticosterone-treated mice. Scale bar (left) = 250 µm, scale bar (right) = 50 µm. **P.** The number of dVenus$^+$ neurons in the DG was increased in corticosterone-treated mice 24 h after conditioning. Saline: N=18; CORT: N=14 mice/group. **Q.** Schematic overview of the fear conditioning paradigm. Mice were sacrificed 90 min after exposure to context A or B. **R.** Post-training CORT treatment resulted in an increase in the number of c-fos$^+$ cells upon exposure to context B, but not to context A. Context A: Saline: N=8; CORT: N= 7 mice/group. Context B: Saline: N=8; CORT: N=6 mice/group. **S.** Testing in either context A or context B did not increase the number of dVenus$^+$ compared to untested controls. N=8 mice/group. **T.** Representative image showing c-fos$^+$ and dVenus$^+$ cells after exposure to B.

Scale bar = 50 µm. **U.** Colocalization between dVenus$^+$ and c-fos$^+$ neurons did not differ between saline and corticosterone-treated mice after re-exposure to context A. Saline: n = 8; CORT: n = 6 mice per experimental group. **V.** Compared to control mice, CORT-treated mice displayed increased colocalization of c-fos$^+$ cells in dVenus$^+$ neurons after exposure to context B. Saline: n = 8; CORT: n = 6 mice per experimental group.

Data are means + s.e.m. Statistical analysis was done with Student's unpaired t-test in **C, E, H, J, L, M** and **P**, and a two-way ANOVA followed by a Tukey post hoc test was used in **R, S, U** and **V**. *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$.

Journal Pre-proof

Figure 3. Corticosterone selectively enhanced the excitability of dVenus$^+$ neurons after fear conditioning. A. Representative traces of excitability measurements**. B.** Corticosterone did not alter AP frequency (in Hz) in dVenus$^-$ neurons after training. Saline: n=12; CORT: n=13 neurons from n = 5 mice/experimental group. **C.** Corticosterone enhanced the excitability of dVenus$^+$ neurons. Saline: n=11; CORT: n=17 neurons from n = 5 mice/experimental group. **D.** Corticosterone without training did not alter AP frequency in dVenus$^-$ neurons compared to saline injected mice. Saline: n=18; CORT: n=13 neurons from n = 5 mice/experimental group. **E.** Corticosterone without training does not alter AP frequency in dVenus$^+$ neurons. Saline: n=15; CORT: n=11 neurons from n = 5 mice/experimental group. **F.** Typical examples of mEPSC traces. **G.** The frequency of the mEPSCs was enhanced in dVenus$^+$ neurons compared to dVenus$^-$ cells, irrespective of treatment. Corticosterone increased the mEPSC frequency in dVenus$^-$ neurons. dVenus$^-$: Saline: n=9; CORT: n=8 neurons from n = 4 mice/experimental group. dVenus$^+$: Saline: n=9; CORT: n=8 neurons from n = 4 mice/experimental group. **H.** No effect of treatment was observed on the mEPSC amplitude in dVenus$^-$ and dVenus$^+$ neurons. dVenus$^-$: Saline: n=9; CORT: n=8 neurons from n = 4 mice/experimental group. dVenus$^+$: Saline: n=9; CORT: n=8 neurons from n = 4 mice/experimental group. Data are means + s.e.m. Statistical analysis was done with a repeated measures ANOVA on the combined data in **B** and **C**, and on the combined data in **D** and **E**, and a two-way ANOVA was conducted on the data in **G** and **H**. *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$.

Journal Pre-proof

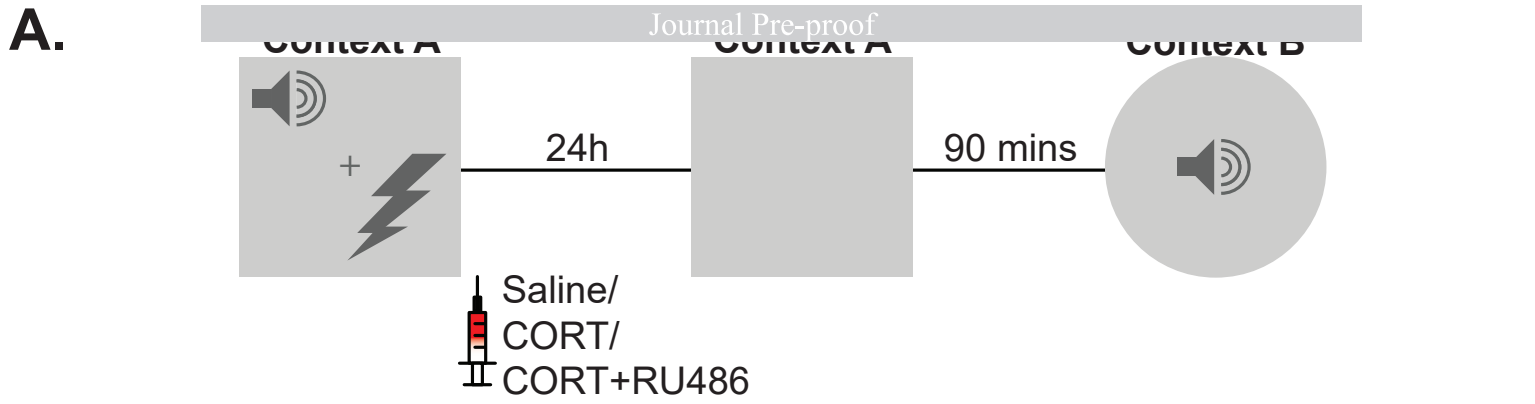
Figure 4. Chemogenetic suppression of the training-activated DG ensemble reduced corticosterone-induced generalized fear.

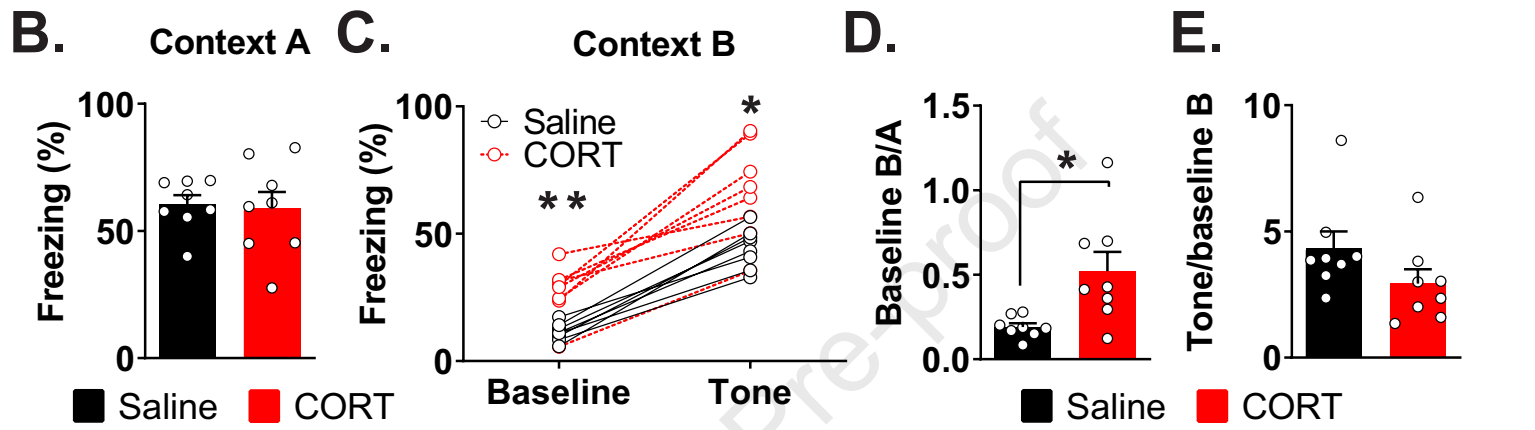
A. Schematic representation of Fos promoter-driven and 4OHT-dependent expression of hM4Di-mCherry in DG neurons. **B.** Timeline of injection, viral transduction and behavioral testing paradigm. Freezing in context A and context B (± tone) was assessed at 24 h and 48 h after training, respectively. **C.** Freezing in context A did not differ between groups. Number of animals mCherry: n = 7 mice; Number of animals hMdDi-mCherry: n = 10 mice. **D.** CNO-induced suppression of hM4Di-mCherry^{+} neurons reduced baseline freezing levels in the neutral context. Freezing to the tone was unaffected. Number of animals mCherry: n = 7 mice; Number of animals hMdDi-mCherry: n = 10 mice. **E.** CNO-induced suppression of hM4Di-mCherry^{+} neurons decreased freezing in context B relative to context A. **F.** Representative image showing the expression of hM4Di-expressing DG neurons and c-fos colocalization. Blue: DAPI, red: mCherry, white: c-fos. Scale bar = 125 µm. **G.** Percentage of c-fos^{+} cells within mCherry^{+} and mCherry^{-} populations of both groups. mCherry: n = 7 mice; hM4Di-mCherry: n = 10 mice. Data are means + s.e.m. Statistical analysis was done with a student's t-test in **C** and **E**, repeated measures ANOVA, followed by Sidak post hoc testing in **D**, and two-way ANOVA followed by Tukey post hoc testing in **G**. *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$.

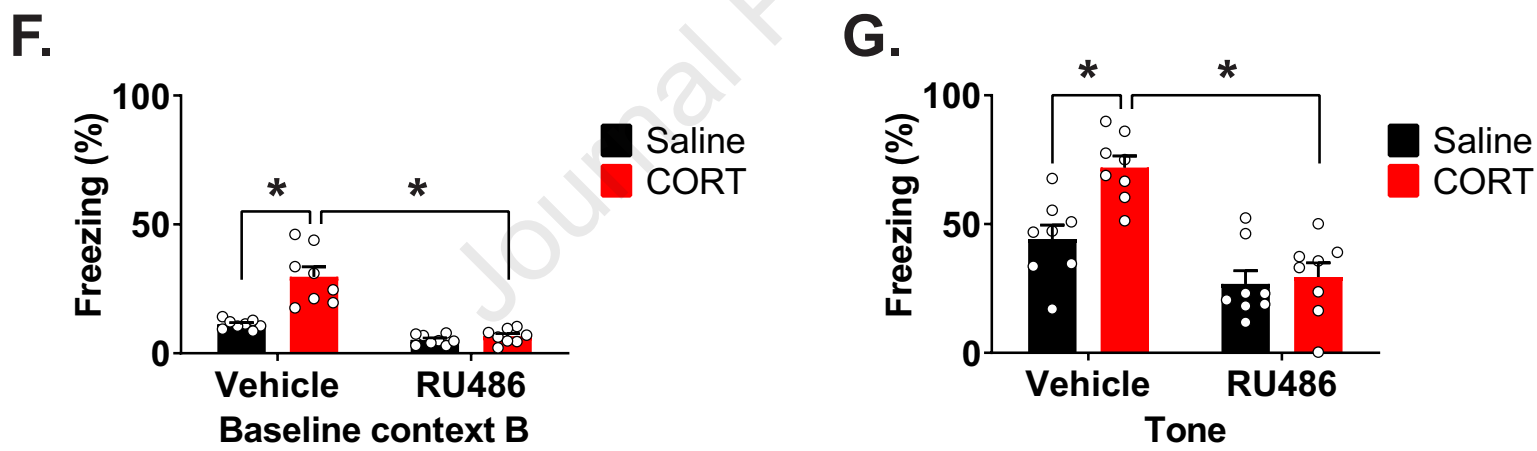
Journal Pre-proof

Journal Pre-proof

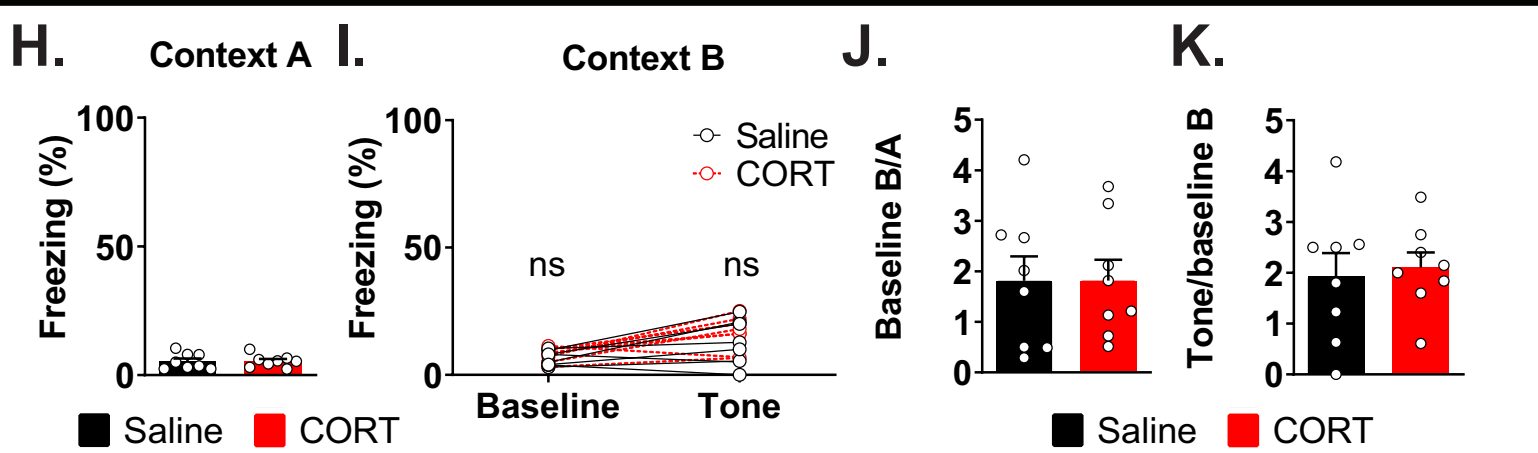
Fear conditioned





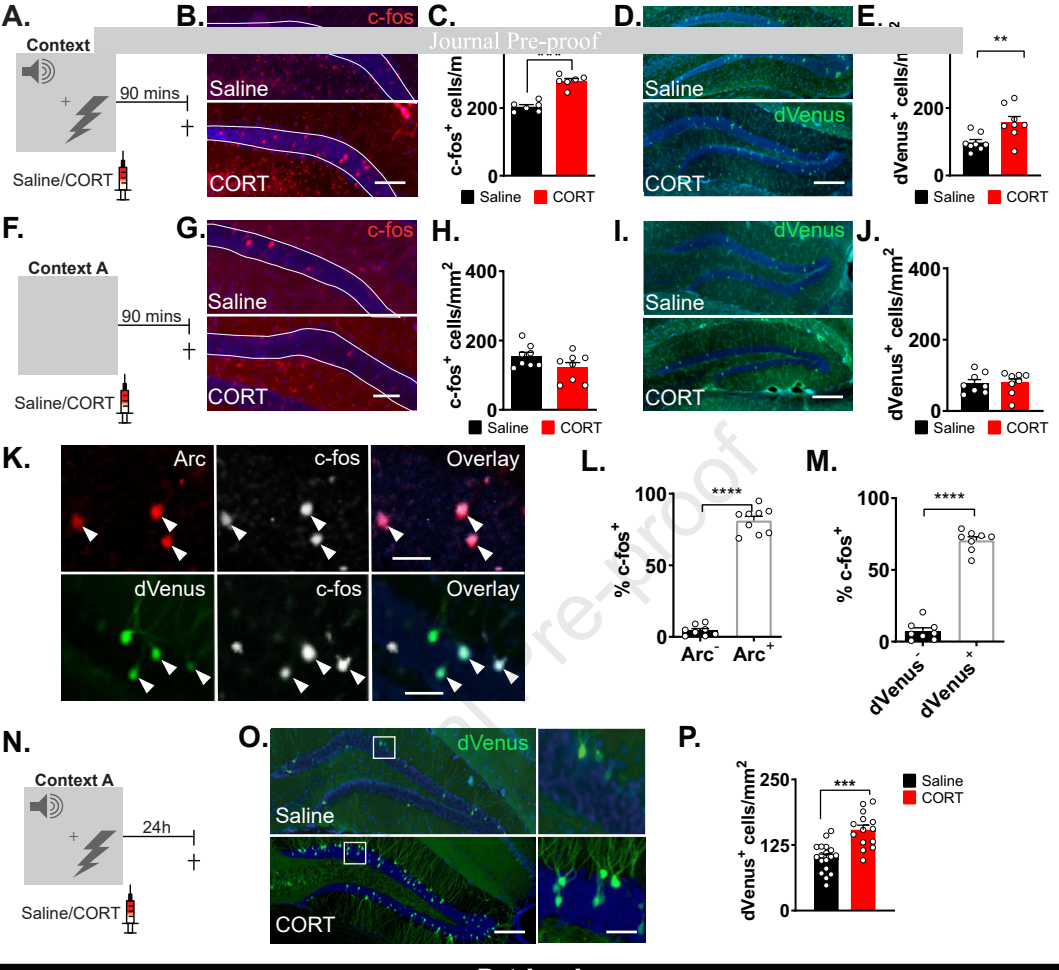


No shock

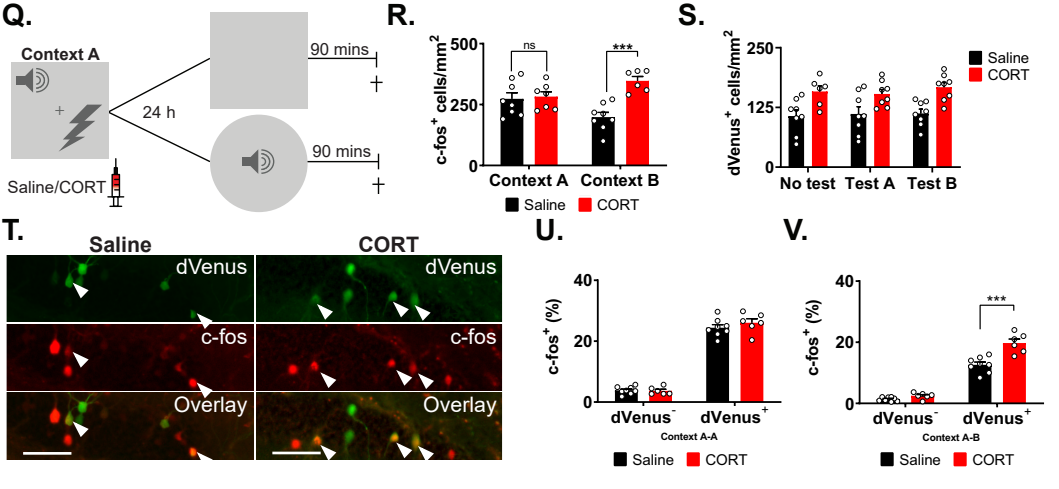


Journal Pre-proof

Consolidation

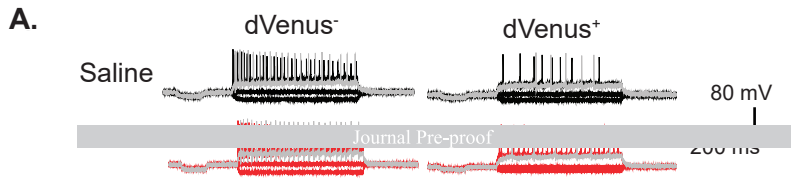


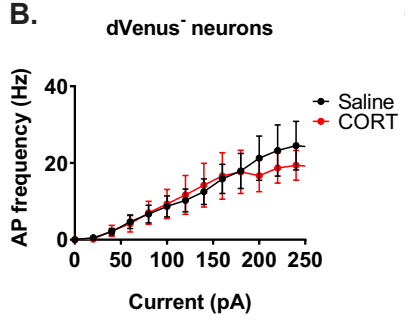
Retrieval

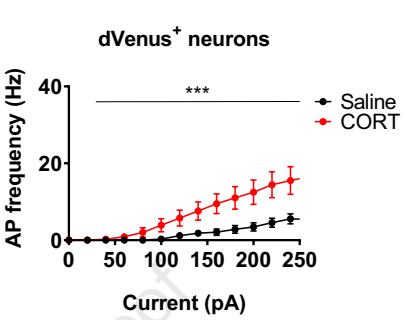


Journal Pre-proof

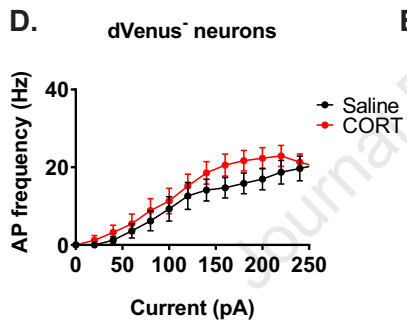
Fear conditioned

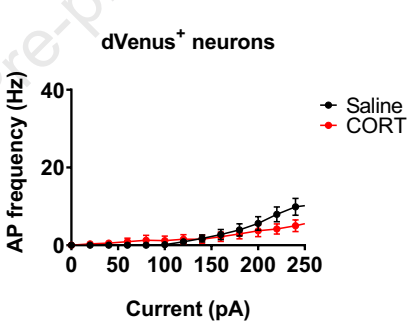




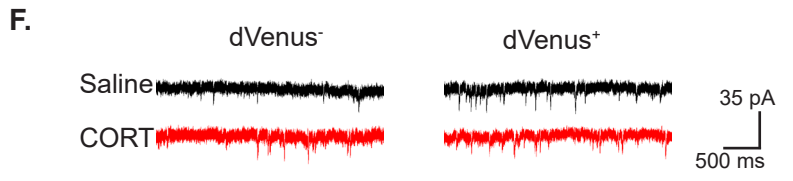


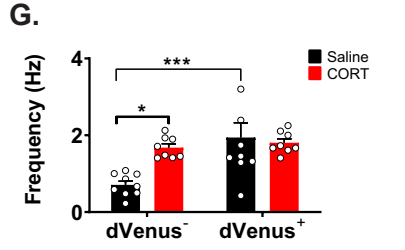
Control

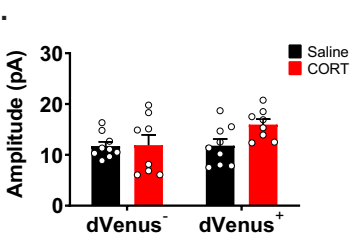




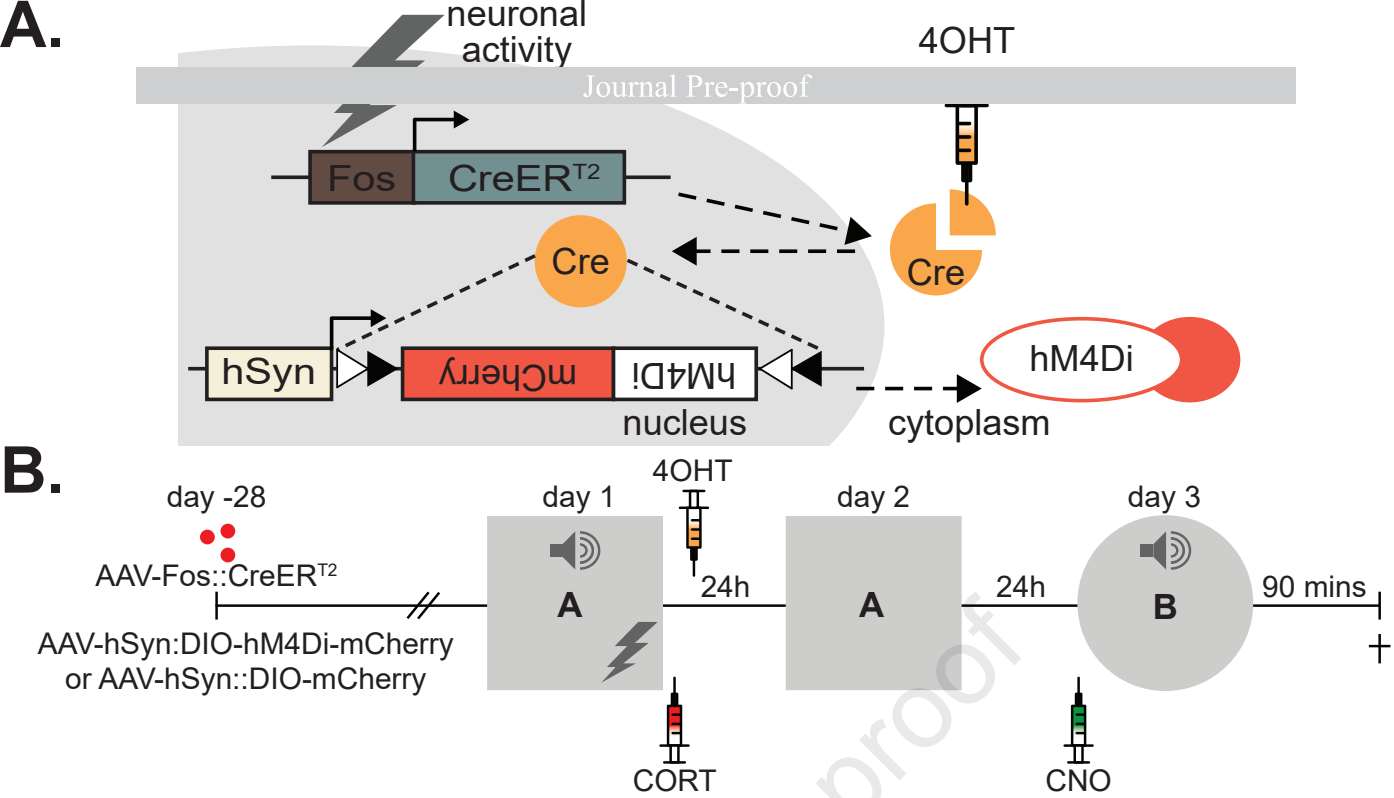
Fear conditioned

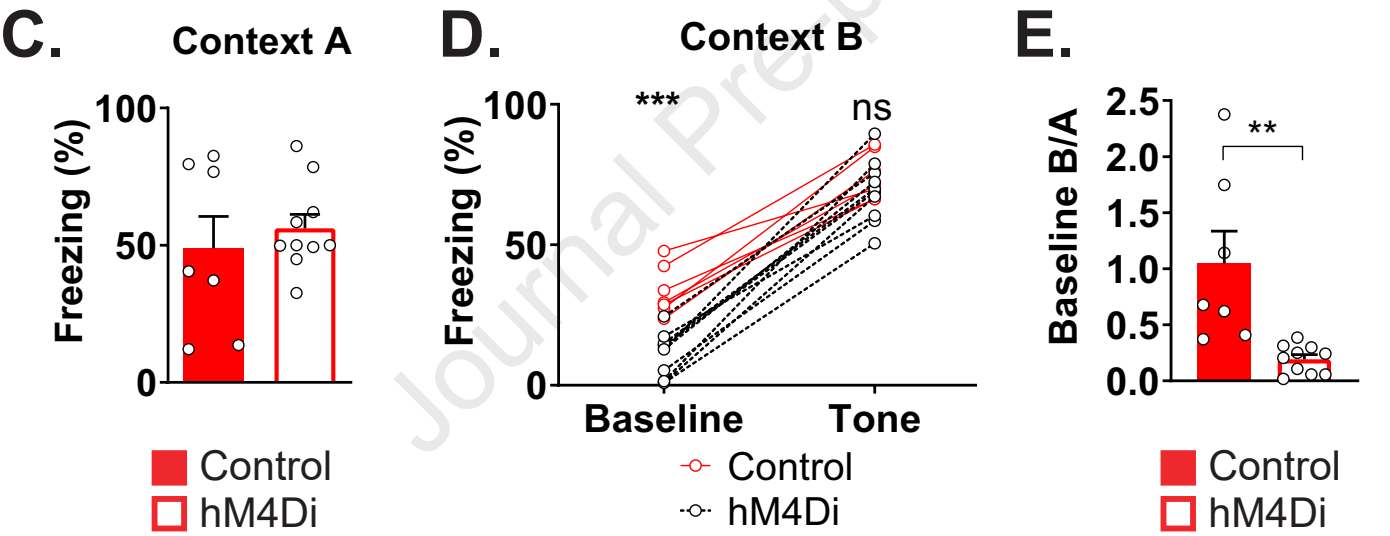


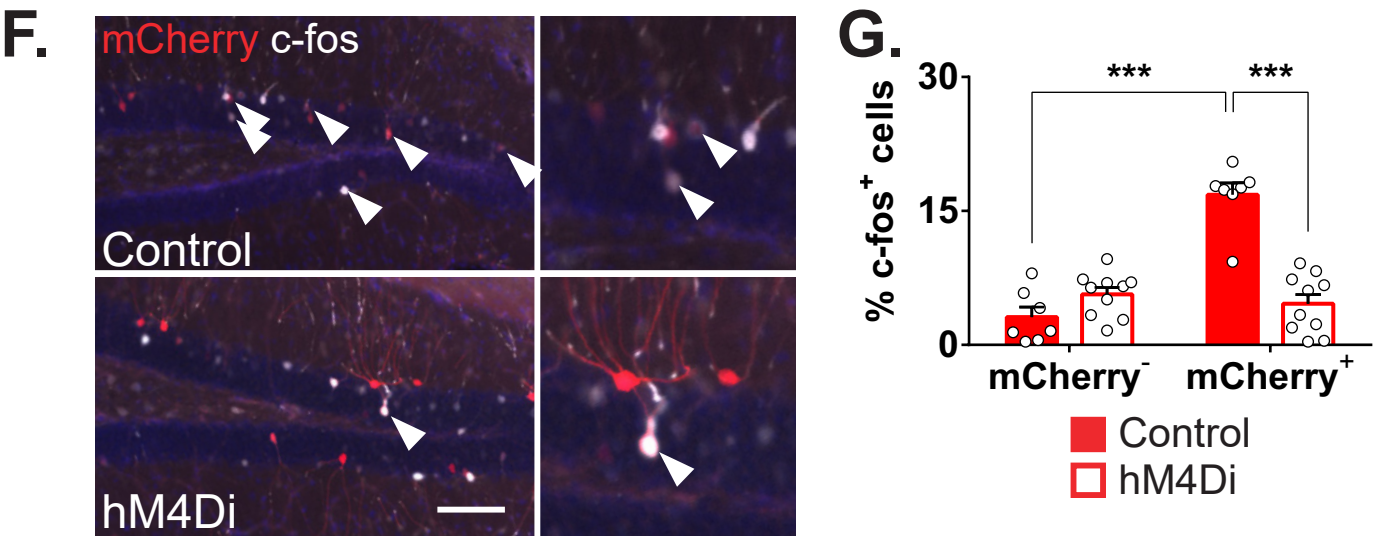




Journal Pre-proof







EJN European Journal of Neuroscience

FENS Federation of European Neuroscience Societies

European Journal of Neuroscience, Vol. 41, pp. 1358–1371, 2015 doi:10.1111/ejn.12897

COGNITIVE NEUROSCIENCE

Differentiating neural systems mediating the acquisition vs. expression of goal-directed and habitual behavioral control



Mimi Liljeholm,[1,2] Simon Dunne[1,3] and John P. O'Doherty[1,3]
[1]Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland
[2]Department of Cognitive Sciences, University of California, 2312 Social & Behavioral Science Gateway Building, Irvine, CA 92697, USA
[3]Division of the Humanities and Social Sciences and Computation and Neural Systems Program, California Institute of Technology, Pasadena, CA, USA

Keywords: action–outcome, functional magnetic resonance imaging, habit formation, human, stimulus–response

Abstract

Considerable behavioral data indicate that operant actions can become habitual, as demonstrated by insensitivity to changes in the action–outcome contingency and in subjective outcome values. Notably, although several studies have investigated the neural substrates of habits, none has clearly differentiated the areas of the human brain that support habit formation from those that implement habitual control. We scanned participants with functional magnetic resonance imaging as they learned and performed an operant task in which the conditional structure of the environment encouraged either goal-directed encoding of the consequences of actions, or a habit-like mapping of actions to antecedent cues. Participants were also scanned during a subsequent assessment of insensitivity to outcome devaluation. We identified dissociable roles of the cerebellum and ventral striatum, across learning and test performance, in behavioral insensitivity to outcome devaluation. We also showed that the inferior parietal lobule (an area previously implicated in several aspects of goal-directed action selection, including the attribution of intent and awareness of agency) predicted sensitivity to outcome devaluation. Finally, we revealed a potential functional homology between the human subgenual cortex and rodent infralimbic cortex in the implementation of habitual control. In summary, our findings suggested a broad systems division, at the cortical and subcortical levels, between brain areas mediating the encoding and expression of action–outcome and stimulus–response associations.

Introduction

Habitual action selection is defined by insensitivity to changes in the causal efficacy with which actions produce rewards and to the current subjective value of those rewards (Balleine & Dickinson, 1998). Neuroscientific research on humans and rodents has demonstrated that the brain areas mediating habitual performance are dissociable from those supporting more deliberate, goal-directed, action selection (Balleine & Dickinson, 1998; Yin *et al.*, 2004, 2005; Valentin *et al.*, 2007; Tricomi *et al.*, 2009). Intriguingly, work in rodents also suggests that distinct neural substrates make specialized contributions to the development vs. deployment of habits (Killcross & Coutureau, 2003). In contrast, in humans there has been no clear differentiation between brain areas that support habit formation and those that implement habitual control. Although a couple of neuroimaging studies have demonstrated increases in neural activity concomitant with the development of habits in a posterior area of the lateral striatum (Tricomi *et al.*, 2009; Wunderlich *et al.*, 2012), the use of overtraining to induce habitual responding in these studies confounds well-trained performance with habitual control. In the current study, in order to discriminate between neural substrates supporting the acquisition vs. expression of habits, we scanned human participants with functional magnetic resonance imaging as they performed a novel instrumental task (see Task section in Materials and methods), designed to rapidly induce habitual responding without the potentially confounding process of overtraining.

Pharmacological disruptions and electrophysiological recordings of the rodent brain have strongly implicated the dorsolateral striatum in the acquisition of habits. Pre-training lesions of the dorsolateral striatum abolish habit formation (Yin *et al.*, 2004), and distinct changes in neuronal activity patterns (Jog *et al.*, 1999), including substantial decreases in firing rates (Carelli *et al.*, 1997; Tang *et al.*, 2007), have been demonstrated in this area across the development of habitual responding. In contrast, the infralimbic region of the prefrontal cortex has been suggested to support an executive control system that facilitates the expression of habits. Post-training muscimol-induced inactivation of the infralimbic cortex disrupts habitual performance (Coutureau & Killcross, 2003; Haddon & Killcross, 2011), and changes in neuronal ensemble activity patterns in this

Correspondence: Mimi Liljeholm, [2]Department of Cognitive Sciences, as above.
E-mail: m.liljeholm@uci.edu

Received 3 October 2014, revised 6 March 2015, accepted 12 March 2015

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

area occur very late in training and closely track the behavioral manifestation of habitual control (Smith & Graybiel, 2013). Further evidence for the involvement of the infralimbic cortex in the expression of habits comes from studies in which optogenetic perturbation of this area disrupts well-ingrained habitual behavior (Smith *et al.*, 2012). Based on these findings in rodents, we hypothesized that human homologs of the dorsolateral striatum and infralimbic cortex would be involved in the formation and expression of habits respectively, such that behavioral insensitivity to outcome devaluation would correlate with neural activity during learning in the dorsal putamen (Carelli & West, 1991; Draganski *et al.*, 2008), but with activity during actual test performance in the subgenual cortex (Ongur & Price, 2000; Ongur *et al.*, 2003).

Materials and methods

Participants

Twenty volunteers (mean age 21.4 ± 2.63 years, range 19–28 years; 11 males) participated in the experiment. Due to technical problems (loss of power to the stimulus computer), one of the subjects was excluded, yielding a total of 19 participants. All participants were healthy and were recruited locally from the city of Dublin, Ireland. The study was approved by the Trinity College Dublin School of Psychology research ethics committee, and all participants gave informed consent. The study conformed to the guidelines set out in the 2013 World Medical Association Declaration of Helsinki.

Task

Goal-directed actions, defined by their sensitivity to changes in both action–outcome contingency and outcome value, have been proposed to depend on an internal model of the world that explicitly relates alternative actions to future environmental states (Doya *et al.*, 2002; Daw *et al.*, 2005). Consistent with this theoretical framework, data from rodent studies suggest that reliance on a goal-directed vs. habitual strategy might depend on the ease with which alternative actions can be associated with distinct outcomes; goal-directed performance appears to dominate, in spite of overtraining, when alternative actions yield distinct sensory-specific outcomes (e.g. grain vs. sucrose pellets) (Colwill & Rescorla, 1985; Holland, 2004), as well as when the rate of outcome delivery depends on the rate of responding, rather than on a particular time interval passing between successive reinforced responses (i.e. ratio vs. interval schedules of reinforcement) (Dickinson *et al.*, 1983). The current task was structured on these potential bases of behavioral control, in that external contingencies either facilitated or impeded a reliable mapping of alternative actions to distinct sensory-specific outcome states.

Participants were required to maintain the balance of a system of fluid-filled beakers (see Fig. 1A for details). As long as all beakers had sufficient fluid, system balance was maintained and randomly occurring balance checks yielded monetary reward. However, on each trial, one of the beakers would be emptied causing 'system imbalance', with balance checks resulting in monetary loss until the participant refilled the beaker by performing a particular instrumental action. The emptying of a beaker was always accompanied by the onset of one of four abstract cues. In the Stimulus–Response (S-R) condition, the identity of the presented cue determined which instrumental action would refill the emptied beaker, regardless of which of the beakers had lost its fluid. Consequently, across trials, each rewarded action was paired with a specific antecedent cue, but was decorrelated from the refilling of any particular beaker. Conversely, in a Response–Outcome (R-O) condition, each instrumental action refilled a particular beaker, regardless of which abstract cue was presented, such that identification of the relevant subgoal (e.g. refilling beaker 1), combined with knowledge about specific action–outcome contingencies (i.e. action 1 refills beaker 1), indicated which action would restore system balance. To ensure that discriminatory neural activity was not due to differences in the visual processing of abstract cues vs. beakers, a matching task was interleaved with the instrumental task (see Fig. 1B).

The persistent execution of an action after its outcome has been devalued is a defining feature of habitual performance (Adams & Dickinson, 1981; Adams, 1982). In the current study, following acquisition of the instrumental task, we devalued one of the four beakers by degrading its relationship to monetary gain, such that system balance was maintained, and continued to yield points even when the liquid in this beaker dropped below threshold. Because there was a small cost for regulating the system, attempts to refill this beaker now resulted in a net loss (see Fig. 1A and Experimental Procedure section for details). Based on the notion that failure to associate alternative actions with distinct outcome states obstructs goal-directed encoding, we predicted that the S-R condition would bolster habit formation during acquisition and bias participants towards habitual action selection, defined as responding to refill the devalued beaker, in subsequent test performance. Note that, in both conditions, distinct features of the stimulus environment can enter into S-R associations. Whereas, in the S-R condition, each rewarded action was reliably preceded by a particular arbitrary cue, rewarded actions in the R-O condition were reliably preceded by the emptying of a particular beaker. However, critically, it was only in the R-O condition that alternative actions could be associated with distinct outcome states. Consequently, we expected performance in this condition to be goal-directed.

Experimental procedure

We scanned participants as they acquired and performed the instrumental task, as well as during a subsequent devaluation test phase. Each subject participated in both the R-O and S-R condition, with conditions being run in separate, immediately consecutive, sessions (order counterbalanced across subjects) and with a novel set of four instrumental actions being used in each condition. Each condition included a response pre-training phase, learning phase, devaluation phase, and final test phase as described below. The response-training phase of the first condition was conducted outside the scanner (in a separate testing room), before participants were transferred to the scanner in which they remained throughout all subsequent stages of the experiment. Before being transferred to the scanner, participants were presented with a cover story describing the beaker system and task. They were told that they would be in one of two possible conditions (one in which each instrumental action refilled a particular beaker regardless of which cue was presented, and another in which the identity of the cue determined which of the four actions was required to refill an emptied beaker, regardless of the identity of that beaker) and that part of their task was to determine which of the two conditions they were in. The entire experiment lasted for approximately 2 h, with 1.5 h being spent in the scanner, and with approximately 60 min of active scanning during the learning phases and devaluation tests in each condition.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

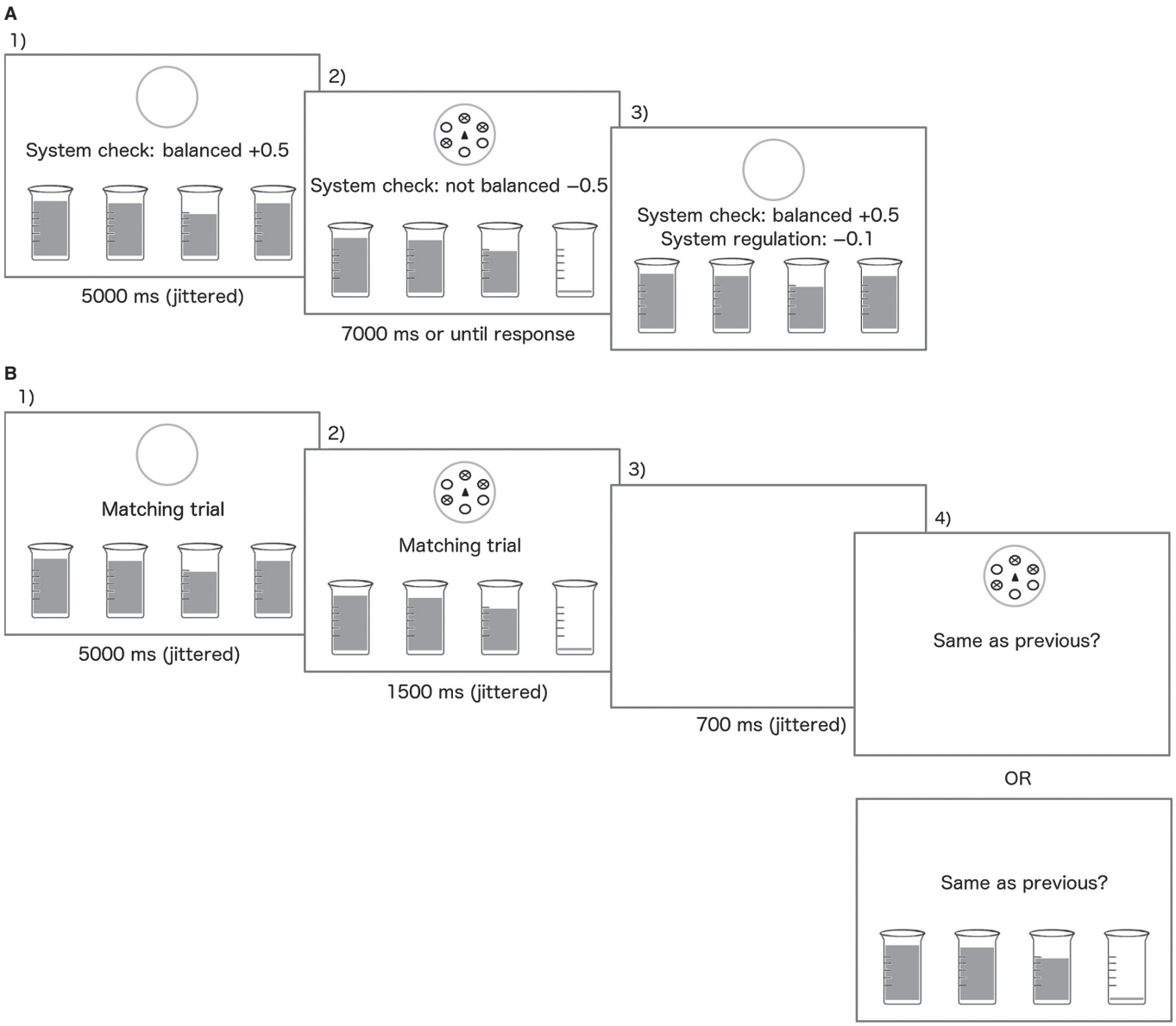


FIG. 1. Illustrations of instrumental learning and matching control tasks. (A) Instrumental learning. Participants are required to maintain the balance of a system of fluid-filled beakers using a set of four instrumental actions (each a three-press sequence, see Materials and methods). During the inter-trial interval, the liquid in the beakers continually fluctuates but remains high, and 'balance checks', occurring at brief random intervals, yield points for system balance (1). At the trial onset, one of four abstract cues appears, the liquid in one of the beakers drops to its bottom, and balance checks begin to indicate a loss of points due to system imbalance (2). Points are continually lost until the participant successfully refills the emptied beaker using one of the four actions. Following completion of the correct action (3), the abstract cue disappears, the beaker is refilled, a small fee is charged for regulating the system, and balance checks again yield points for system balance. If the correct action is not performed within 7 s, the beaker is automatically refilled, in which case there is no charge for system regulation. (B) Matching task. The inter-trial interval (1) and trial onset (2) were as in the instrumental task but were followed, at 1500 ms after trial onset, by a blank screen for 700 ms (3). The subsequent, final screen (4) showed a matching/non-matching stimulus together with a query about the match. In the S-R condition, the final screen always showed one of the abstract cues (top); conversely, in the R-O condition, the stimulus to be matched was always a set of beakers (bottom).

Response pre-training

Prior to the instrumental learning phases, participants received pre-training on the four instrumental actions (each being a three-press sequence). During this training, key-press sequences were illustrated by a white dot moving across three gray squares, horizontally aligned at the center of the screen. Initially, participants viewed and then immediately attempted to replicate each sequence, with feedback (i.e. correct/incorrect) given on each trial. After a total of five correct replications of each response sequence, they proceeded to a retrieval phase, in which they had to generate each unique sequence at least five times without any prompts, again with feedback given at the completion of each sequence. Participants were allowed to repeat these two phases as many times as they wanted to, knowing that they would have to use the actions to earn monetary reward in a subsequent phase.

Instrumental learning phase

The instrumental task was as illustrated in Fig. 1A. Note that, in addition to the increase or decrease in monetary points based on

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

system balance, there was a small cost for regulating the system. This response cost was included to ensure that, during test, participants would not respond simply based on any reinforcement intrinsic to executing the correct response. The response cost message (screen 3 in Fig. 1A) also served to inform participants that their action had successfully regulated the system, rather than the system having self-regulated (in which case no response cost was charged) due to failure to perform the correct action within 7 s of trial onset. Participants were allowed to perform as many key-presses as they wanted during system imbalance (i.e. during the 7 s following trial onset), with the correct sequence of three consecutive key-presses immediately restoring system balance and terminating the trial. There was no constraint on the temporal spacing between key-presses, as long as all three presses were performed consecutively within the trial window.

Critically, the stimulus materials presented in Fig. 1A were identical across the two conditions; our manipulation consisted entirely of differences in the contingencies between cues, actions and beaker outcomes. To rule out visual processes involved in selectively attending to the abstract cues vs. the beakers as a source of any imaging effects, a matching task was block-interleaved with the instrumental task during instrumental learning (see Fig. 1B). Briefly, in matching blocks, the inter-trial intervals and trial onsets were exactly as in the system balance task, except that the words 'Matching trial' were displayed center screen. Without this indication, matching trials would be identical to instrumental trials during the relevant trial period, and thus could not have served as controls. Following the appearance of the abstract cue and emptying of the relevant beaker, a white masking screen was displayed, followed by a depiction of either an abstract cue (S-R condition) or a set of beakers (R-O condition), together with a query about whether the currently shown cue/beaker set matched that on the previous screen. In each condition, the instrumental learning phase consisted of four blocks of trials, with each block being further divided into one sub-block of 24 instrumental trials, followed by a sub-block of eight matching trials, and with the order of trials randomized within each sub-block.

At the end of the instrumental learning phase, participants were asked whether they felt confident that they had learned how to regulate the system or whether they wanted to receive an additional set of 20 training trials (five with each action). Five participants (two in the R-O and three in the S-R condition) requested and received additional training. No scanning was conducted during additional training, nor were the added trials included in assessments of accuracy and response times during acquisition.

Devaluation phase

Following instrumental learning, participants were instructed that the system had changed such that one of the beakers was no longer relevant for system balance, which would be maintained, and continued to yield points even when the liquid in this beaker dropped below threshold. They then observed as the system regulated itself (i.e. no actions were performed) across 16 trials (four with each beaker) in order to discover the identity of the devalued beaker. Participants were told that they would not lose or gain any of the displayed points during this phase. In the imaging experiment, two participants failed to correctly identify the devalued beaker after this devaluation procedure, which was therefore repeated once for these participants. All other participants successfully identified the devalued beaker at the end of the devaluation procedure.

Test phase

Having correctly identified the devalued beaker, participants were again given the opportunity to regulate the system for personal monetary reward. During this phase, all text messages indicating gains or losses were covered up, in order to prevent additional learning (i.e. simulating extinction). Participants were instructed that, in spite of these gray strips, they should assume that all was exactly as they had learned before, i.e. they would still lose points whenever the system was not balanced, there was still a cost for regulating the system, and the previously identified irrelevant beaker was still irrelevant for system balance. Importantly, because there was a small charge for each instrumental regulation of the system, refilling the now irrelevant beaker resulted in a net monetary loss. The test phase consisted of a single instrumental block with 44 randomly ordered trials, i.e. 11 trials with each beaker, including the devalued (i.e. irrelevant) one. We operationally defined devaluation insensitivity, our assay of habitual performance, as the proportion of the 11 devalued trials on which participants initiated a response to refill the beaker. Thus, if a participant initiated a response on two of those 11 trials, their devaluation insensitivity score would be 0.18.

Pay-off structure

Throughout the instrumental task, system balance checks occurred on average every 3 s. During the inter-trial interval (mean 5 s), the system was always balanced, with each check yielding a reward of 0.5 points. At each trial onset, the system became imbalanced, with each balance check resulting in a loss of 0.5 points, and would remain so for 7 s unless regulated by the participant. Thus, if no action were taken to balance the system, each trial would entail an average loss of 1.17 points. If the system was balanced immediately, no points were lost due to system imbalance, but there was a response cost of 0.1 points (this cost was only applied to correct responses resulting in system regulation). Thus, the average gain of responding to balance the system was 1.07 points. After beaker devaluation, during the test phase, this was still the case for regulatory responses refilling non-devalued beakers; however, responding to refill the devalued beaker now resulted in a net loss of 0.1 points (the response cost).

Imaging acquisition and analyses

A 3 Tesla scanner (MAGNETOM Trio, Siemens) was used to acquire structural T1-weighted images and T2*-weighted echoplanar images (repetition time, 2.65 s; echo time, 30 ms; flip angle, 90°; 45 transverse slices; matrix, 64 × 64; field of view, 192 mm; thickness, 3 mm; slice gap, 0 mm) with blood oxygenation level-dependent contrast. To recover signal loss from dropout in the medial orbitofrontal cortex (O'Doherty *et al.*, 2002), each horizontal section was acquired at 30° to the anterior commissure–posterior commissure axis. Image processing and statistical analyses were performed using statistical parametric mapping (SPM)8 (http://www.fil.ion.ucl.ac.uk/spm). The first four volumes of images were discarded to avoid T1 equilibrium effects. All remaining volumes were corrected for differences in slice acquisition, realigned to the first volume, spatially normalized to the Montreal Neurological Institute echoplanar imaging template, and spatially smoothed with a Gaussian kernel (8 mm, full width at half-maximum). We used a high-pass filter with a cutoff of 128 s.

As previously noted, instrumental conditions were run in separate consecutive sessions, with the acquisition phase of each condition

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

consisting of interleaved instrumental and matching block, and with active scanning occurring only during the acquisition and test phase of each condition (i.e. the scanner was off during response pre-training and devaluation phases), for a total of four active scanning periods per subject. For each subject, we constructed a design matrix that included the acquisition and test phases of both experimental conditions. For each of the two instrumental acquisition phases (i.e. R-O and S-R), stick functions modeled the trial onsets in each block of instrumental learning and in matching blocks. In addition, three regressors, respectively modeling the onsets of error trials, the times of all individual key-presses and the times of all point displays (i.e. gains and losses), were entered as regressors of no interest, together with six regressors accounting for the residual effects of head motion. For each of the two test phases, two stick functions respectively modeled non-devalued and devalued trials. Again, for each instrumental condition, additional regressors modeling the onsets of error trials and the times of all key-presses were added together with six regressors accounting for the residual effects of head motion as regressors of no interest. All regressors were convolved with a canonical hemodynamic response function.

Group-level random-effects statistics were generated by entering contrasts of parameter estimates for the different regressors into between-subjects analyses of variance (ANOVAS). Specifically, to delineate neural substrates engaged during early acquisition vs. during expression, contrasts of parameter estimates for the first two blocks of instrumental learning, and for the devaluation test phase, were entered for each instrumental condition into a 2 × 2 (condition by experimental phase) ANOVA. Contrasts of parameter estimates for all learning blocks and for matching blocks were entered into a separate 2 × 2 (stimulus by task) ANOVA to compare differences between instrumental conditions during learning with those between matching control conditions. An analogous ANOVA was performed using contrasts of parameter estimates for devaluation test trials and matching trials. Finally, a 4 × 2 (acquisition block by instrumental condition) ANOVA assessed training-related changes in neural activity, with contrasts of parameter estimates for each of the four blocks of instrumental learning, for each instrumental condition. Following estimation of the second-level model, F-tests were specified by adding linear weights to each block of learning (e.g. [−1.5 −0.5 0.5 1.5] for increases across blocks and [1.5 0.5 −0.5 −1.5] for decreases) in each instrumental condition.

To assess whether neural discrimination between instrumental conditions correlated with the degree of devaluation insensitivity, a simple *t*-test was performed on first-level interaction contrasts [S-R > R-O × Instrumental > Matching] of parameter estimates from the learning phase, with the degree of devaluation insensitivity entered as a covariate. An analogous *t*-test using interaction contrasts of parameter estimates from the devaluation test phase was also performed, and exclusive functional masks were used to assess the specificity of neural effects, such that all voxels that reached significance at a threshold of $P < 0.1$ when assessing correlates of devaluation insensitivity during the test phase were removed from the effects observed during learning, and vice versa. Exclusive functional masks were also used to assess the directions of simple effects underlying the Instrumental Condition by Experimental Phase interaction. Finally, we used exclusive functional masks to selectively assess training-related increases vs. decreases in neural activity, such that, for example, when testing for increases in neural activity across blocks of training in the R-O condition, voxels were removed that reached significance at $P < 0.1$ for the same test in the S-R condition.

Small volume corrections (SVCs) were performed on several *a-priori* regions of interest using a 10 mm sphere, with center coordinates obtained from highly relevant studies. Specifically, in an analogous study on observational learning (Liljeholm *et al.*, 2012), we found selective recruitment of the extrastriate cortex (±45, −72, 9), the tail of the caudate nucleus/thalamus (±21, −30, 9) and the lingual gyrus (±12, −69, 0) by the S-R condition during acquisition. Conversely, areas selectively recruited by the R-O condition in that same observational learning study, and also identified in our other work on goal-directed performance (Liljeholm *et al.*, 2011), include the inferior parietal lobule (IPL) (±51, −52, 33) and anterior caudate nucleus (±16, 8, 19). Finally, several human neuroimaging studies have implicated a posterior region of the putamen (±33, −24, 0) in habit formation (Tricomi *et al.*, 2009; de Wit *et al.*, 2012; Wunderlich *et al.*, 2012; Lee *et al.*, 2014), and the ventromedial prefrontal cortex (±4, 55, −7) in goal-directed processes (Hampton *et al.*, 2006; Tanaka *et al.*, 2008; Liljeholm *et al.*, 2011). It should be noted that, although specified *a priori* based on a closely related literature, effects in several of these regions survived whole-brain cluster-size thresholding (CST) correction, as well as SVC, in the current study.

As noted, the putamen, considered a human homolog of the rodent dorsolateral striatum based on its afferent and efferent projections (Carelli & West, 1991; Draganski *et al.*, 2008), has been implicated in habitual performance in several human imaging studies (Tricomi *et al.*, 2009; de Wit *et al.*, 2012; Wunderlich *et al.*, 2012; Lee *et al.*, 2014). In contrast, the role of the infralimbic cortex in instrumental performance has been exclusively demonstrated in rodents (Killcross & Coutureau, 2003; Haddon & Killcross, 2011; Smith *et al.*, 2012; Smith & Graybiel, 2013). Consequently, and based on anatomical and functional evidence for a homology between the rodent infralimbic cortex and the human subgenual cortex (Ongur & Price, 2000; Ongur *et al.*, 2003; Drevets *et al.*, 2008), predictions regarding this area were tested using an anatomical mask of the subgenual cortex (Brodmann area 25), defined using the Wake Forest University Pickatlas (Maldjian *et al.*, 2003). All other effects were reported at $P < 0.05$, using CST of SPM t-maps to adjust for multiple comparisons (Forman *et al.*, 1995). AlphaSim (Ward, 2000), a Monte Carlo simulation, was used to determine cluster size and significance. Using an individual voxel probability threshold of $P = 0.005$ indicated that using a minimum cluster size of 111 Montreal Neurological Institute transformed voxels resulted in an overall significance of $P < 0.05$.

To eliminate non-independence bias for plots of contrast values, a leave-one-subject-out (Esterman *et al.*, 2010) approach was used, in which 19 general linear models were run with one subject left out in each, and with each general linear model defining the voxel cluster for the left-out subject. Using rfxplot (Glascher, 2009), mean contrast values were extracted from spheres (10 mm) centered on these leave-one-subject-out peaks (identified within regions of interest for SVCs) and were averaged across subjects to plot overall effect sizes.

Results

Behavioral results

The results from the test phase indicated that our manipulation did indeed produce differences in devaluation sensitivity. Having correctly identified the devalued beaker, participants initiated a response on a significantly greater proportion of devalued trials in the S-R condition (mean proportion 0.43, SEM 0.09) than in the R-O

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

condition (mean proportion 0.19, SEM 0.09) [$t_{18} = 2.3$, $P < 0.031$]. Indeed, whereas only five of 19 participants initiated a response on any devalued trials in the R-O condition, 15 of 19 participants initiated a response on at least one devalued trial in the S-R condition ($\chi^2 = 10.56$, $P < 0.005$).

Note that devaluation insensitivity was defined such that a response to obtain the devalued outcome (i.e. to refill the devalued beaker) was counted even if the entire three-press sequence was not completed on that trial. Indeed, in the S-R condition, when a participant responded to fill up the devalued beaker, they often did so without completing the entire three-press sequence, consistent with evidence from the rodent literature that the response most proximal to reward remains sensitive to devaluation long after more distal responses have become habitual (Killcross & Coutureau, 2003). In contrast, those few participants that responded on devalued trials in the R-O condition did so on most or all devalued trials and completed all three key-presses whenever initiating a response, suggesting a more deliberate decision to respond.

Importantly, during the test phase, whereas in the R-O condition participants could determine both the accuracy and value of a particular action solely by identifying the emptied beaker, in the S-R condition, determining the value and accuracy of a given response required identification of both the abstract cue and the emptied beaker. It is possible therefore that the differences in devaluation performance were due to this additional aspect of the S-R condition. The overall pattern of behavioral results, however, strongly suggests that this was not the case, and generally rules out task difficulty as the source of our effects. First, there were no significant differences between conditions in the percent of incorrect responses, during either acquisition [R-O: mean 14%, SEM 3; S-R: mean 11%, SEM 2; $t_{18} = 1.1$, $P = 0.30$] or test performance [R-O: mean 4%, SEM 1; S-R: mean 4%, SEM 2; $t_{18} = 0$, $P = 1.0$]. Likewise, response times did not differ between conditions during either acquisition [R-O: mean 1482 ms, SEM 62; S-R: mean 1393 ms, SEM 78; $t_{18} = 1.95$, $P = 0.21$] or test performance [R-O: mean 1258 ms, SEM 54; S-R: mean 1361 ms, SEM 68; $t_{18} = -1.52$, $P = 0.15$]. The fact that there were no significant differences between conditions in either accuracy or reaction times during test makes it highly unlikely that differences in devaluation sensitivity reflected differences in task difficulty. Perhaps most pertinently, differences in devaluation insensitivity were not correlated with differences in accuracy ($P = 0.5$) or reaction times ($P = 0.8$), nor did individual differences in accuracy or reaction times predict neural effects in any of the areas identified by our imaging analyses.

To assess the influence of counterbalancing order, we performed order-by-condition analyses of variance on response times and accuracy scores during the learning and test phases, as well as on the devaluation insensitivity scores. There was no significant interaction between order and condition for devaluation insensitivity ($F_{1,17} = 0.55$, $P = 0.47$), or for either response times ($F_{1,17} = 0.89$, $P = 0.36$) or accuracy scores ($F_{1,17} = 0.07$, $P = 0.79$) during the test phase. In contrast, during training, there was an anticipated interaction for both response times ($F_{1,17} = 5.96$, $P < 0.03$) and accuracy scores ($F_{1,17} = 23.01$, $P < 0.001$), such that response times were longer and the percent incorrect was greater for whichever condition came first. Thus, whereas the difference in response times between instrumental conditions differed across the two orders (mean difference for R-O first 0.19; mean difference for S-R first −0.32), there was no difference between instrumental conditions when compared for a given order (i.e. for R-O vs. S-R with both first, $P = 0.84$ and for R-O vs. S-R with both second, $P = 0.71$). A similar pattern was observed for percent incorrect scores (mean difference for R-O first 0.08; mean difference for S-R first −0.01; significance test for R-O vs. S-R with both first, $P = 0.26$ and for R-O vs. S-R with both second, $P = 0.90$). Consequently, we can rule out counterbalancing order as a source of our behavioral and imaging effects of interest.

There were no differences between instrumental conditions in the number of requested repetitions of response pre-training phases [R-O: mean 1.98, SEM 0.29; S-R: mean 1.75, SEM 0.18; $t_{18} = 0.72$, $P = 0.48$], or in the total number of key-presses executed during instrumental learning [R-O: mean 470.63, SEM 32.14; S-R: mean 503.74, SEM 31.63; $t_{18} = -0.85$, $P = 0.40$] or on non-devalued trials during test [R-O: mean 121.84, SEM 4.57; S-R mean: 125.53, SEM 4.53; $t_{18} = -1.34$, $P = 0.57$]. Finally, there was no difference between instrumental conditions in the total number of points earned during instrumental learning [R-O: mean 33.6, SEM 3.61; S-R: mean 33.4, SEM 3.33; $t_{18} = 0.071$, $P = 0.94$]. This allowed us to rule out the number of motor responses as well as total earnings as sources of the difference in devaluation performance.

Neuroimaging results

Our primary objective was to investigate whether neural discrimination between the two instrumental conditions differed across the acquisition and expression of behavioral control, particularly with respect to blood oxygenation level-dependent activity predictive of individual differences in devaluation insensitivity (see Table 1). To rule out the possibility that differences between our instrumental conditions were due to differences in the processing of relevant visual features (i.e. of abstract cues in the S-R condition and of the set of beakers in the R-O condition), additional analyses formally contrasted such differences with those emerging between matching control conditions (see Table 2). Unless otherwise noted, all effects reported below survived the subtraction of matching control conditions. All of the results described below survived correction for multiple comparisons at $P < 0.05$ using either whole-brain CST or SVCs based on coordinates from relevant previous studies. Notably, many of the areas specified *a priori* as targets of SVC, including the lingual gyrus, extrastriate cortex and IPL, also survived correction using whole-brain CST, as indicated in the fourth columns of Tables 1 and 2.

Discrimination between instrumental conditions during acquisition vs. performance

To delineate the neural substrates engaged during early acquisition vs. during expression of operant behavior, we entered the imaging data from the first two blocks of instrumental learning together with that from the subsequent test phase, for each instrumental condition, into a 2 × 2 (condition by experimental phase) ANOVA (see Table 1). Although processes supporting acquisition should certainly dominate during the first two blocks of instrumental learning, we cannot completely rule out the possibility that some element of expression was also present during this phase. Inclusion of only the very earliest learning trials would not eliminate this possibility, but would introduce dramatically different error levels and severely reduce the statistical power. We note, however, that the presence of processes related to expression during the early stages of instrumental learning should reduce, rather than enhance, discrimination between the levels of the 'experimental phase' variable. Thus, in areas where we did observe a significant difference in discriminatory activity across the different phases of the experiment, such differences were probably attributable to the differential effects of acquisition vs. expression.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

TABLE 1. Imaging results from a 2 × 2 ANOVA contrasting instrumental conditions and training phases

Test	Area	x, y, z	T at peak level	Correction*	Cluster size at $P < 0.005$*
Main effects					
S-R > R-O	Inferior occipital	36, −85, −8	6.33	CST	859
	Middle occipital	36, −85, 7	6.82		
	Superior occipital	27, −67, 28	5.16		
	SPL	18, −61, 52	4.22		
R-O > S-R	DMPFC	18, 56, 25	3.62	CST	843
	VMPFC	3, 41, 1	3.25	CST	
	IPL	60, −55, 28	3.99	CST	187
	Posterior cing.	−3, −34, 43	3.13	CST	190
	Insula	39, 14, −41	3.89	CST	157
Acq. > Perf.	VMPFC	−3, 62, 4	6.12	CST	2181
	DMPFC	−3, 47, 37	5.65		
	VS	−6, 5, −8	4.90		
	aCN	9, 20, 7	3.53		
	VLPFC	−45, 32, −11	5.46	CST	354
	IPL	−51, −64, 34	4.76	CST	200
Perf. > Acq.	Cerebellum	27, −61, −26	9.17	CST	9773
	SMA	0, 8, 46	5.15		
	Insula/putamen	−33, 8, 4	5.69		
	Post-central S.	−42, −37, 58	6.33		
	Pre-central S.	−30, −4, 52	5.97		
	Calcarine	24, −58, 10	4.94		
	SPL	−15, −64, 49	5.31		
Interactions					
S-R > R-O (Perf.>Acq.)	Occipital	33, −79, 19	5.10	CST	859
	SPL	24, −52, 46	3.58		
R-O > S-R (Perf.>Acq.)	DMPFC	−9, −41, 40	4.26	CST	168

*The absence of an entry in the 'correction' and 'cluster size' columns indicates that the relevant cluster is continuous (at the $P < 0.005$ threshold) with that listed above it. Acq., Acquisition; Perf., Performance; SPL, superior parietal lobule; DMPFC, dorsomedial prefrontal cortex; aCN, anterior caudate nucleus; cing., cingulate; VLPFC, ventrolateral prefrontal cortex; SMA, supplementary motor area; S., post-/pre-central sulcus.

TABLE 2. Imaging results from 2 × 2 ANOVAs contrasting instrumental and matching tasks, and from tests of neural correlates of devaluation insensitivity

Contrast	Area	x, y, z	T at peak level	Correction*	Cluster size at $P < 0.005$*
Learning phase					
S-R > R-O (instr. > ctrl)	EBA	39, −79, 10	4.09	SVC	45
R-O > S-R (instr. > ctrl)	IPL	57, −58, 31	4.41	CST	275
	Insula	33, 14, 4	3.57	CST	421
	Precuneus	6, −70, 37	4.28	CST	395
	Post. cing.	9, −31, 43	3.36		
	aCN	12, 2, 16	3.40	SVC	18
Test phase					
S-R > R-O (instr. > ctrl)	Occipital	30, −79, 19	5.30	CST	1153
	SPL	24, −62, 49	4.47		
R-O > S-R (instr. > ctrl)	IPL	57, −46, 37	3.09	SVC	18
	Insula	42, −1, −5	4.06	CST	383
	DMPFC	−6, 41, 40	3.87	CST	382
Matching effects					
S-R control > R-O control	Lateral LG	24, −91, −8	5.20	CST	198
R-O control > S-R control	Medial LG	−9, −76, −8	5.15	CST	1541
	Cuneus	−9, −91, 19	5.01		
Correlates of devaluation insensitivity					
Positive correlation	CRBL/LG	−6, −52, −8	5.06	CST	125
	tCN/th	−12, −28, 10	4.19	SVC	21
	Subgenual	−15, 23, −14	5.24	CST	224
	VS	−9, 5, −14	8.47		
Negative correlation	IPL	45, −43, 34	4.62	SVC	54

*The absence of an entry in the 'correction' and 'cluster size' columns indicates that the relevant cluster is continuous (at the $P < 0.005$ threshold) with that listed above it. Instr., instrumental; ctrl, control; Post. cing., posterior cingulate; aCN, anterior caudate nucleus; SPL, superior parietal lobule; DMPFC, dorsomedial prefrontal cortex; tCN/th, tail of the caudate nucleus/thalamus; CRBL, cerebellum; LG, lingual gyrus.

Stimulus–Response-related activity

Activity that was greater in the S-R than the R-O condition (Fig. 2) emerged in the middle and superior occipital cortex, and in the superior parietal lobule. Interaction tests and (exclusively masked) tests of simple effects revealed that activity in the superior parietal lobule and superior occipital cortex was significantly greater in the

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

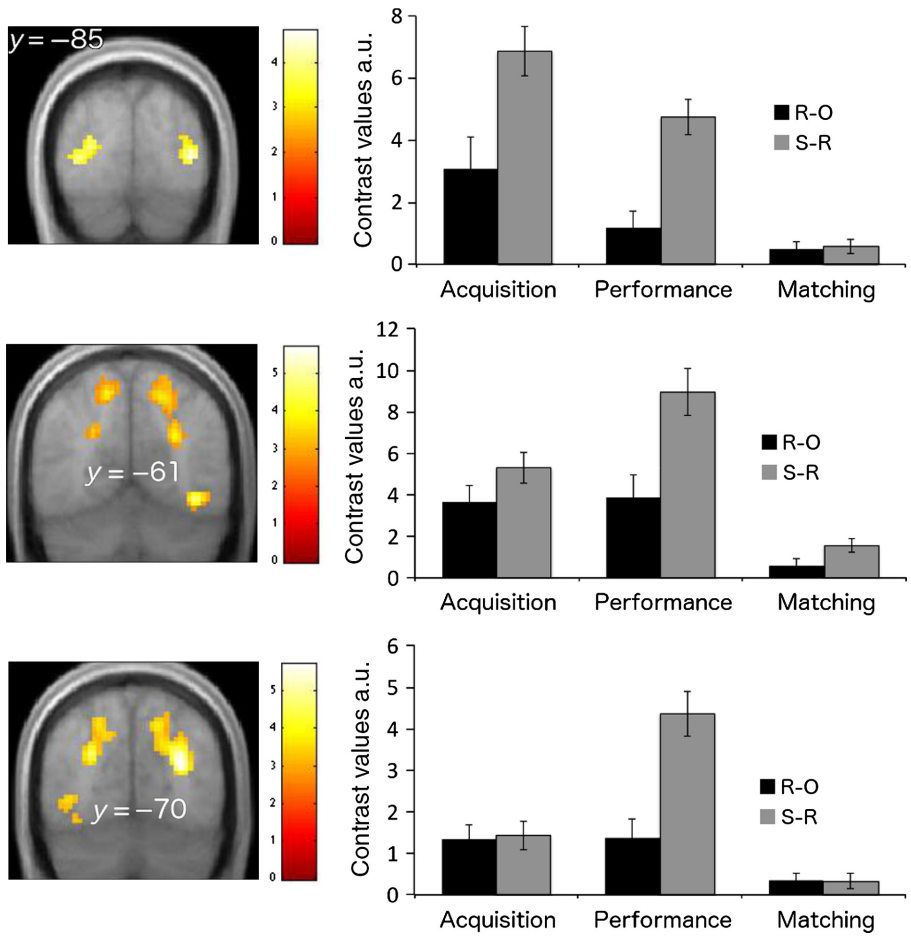


FIG. 2. Areas selectively involved in the S-R condition. Statistical maps show main effects of the S-R > R-O contrast, in the middle occipital cortex (top row), and an interaction effect {i.e. [S-R > R-O (performance > acquisition)]} in the superior occipital cortex, extending into the superior parietal lobule (middle and bottom rows). Bar plots show contrast values estimated at leave-one-subject-out coordinates for each instrumental condition, during both acquisition and performance, and for matching controls. Error bars = SEM. a.u., arbitrary units.

S-R than the R-O condition only during test performance, but not during early acquisition (Table 1). Moreover, during early acquisition, only a small area of the middle occipital cortex, the extrastriate body area (EBA; SVC), survived subtraction of corresponding matching controls (Table 2). In contrast, S-R selective activity throughout the occipital cortex, including the EBA, survived subtraction of matching controls during the test phase.

Response–Outcome-related activity

Activity significantly greater in the R-O condition than the S-R condition (Fig. 3) was observed throughout a fronto-parietal network, including the dorsal and ventral medial prefrontal cortex (VMPFC), IPLs, posterior cingulate and bilateral insula. Interaction tests and (exclusively masked) tests of simple effects revealed that, in the dorsomedial prefrontal cortex, discrimination between instrumental conditions occurred during test performance only (Table 1). Analyses assessing differences between conditions relative to matching controls also revealed R-O selective effects in the dorsal anterior caudate, but did not yield significant effects in the VMFPC (Table 2).

Discrimination between conditions across blocks of acquisition

To further assess differential activity related to acquisition processes, we tested for differences between conditions in training-dependent changes in neural activity, by adding linear weights to blocks of instrumental learning. An interaction test assessing neural activity that increased in the R-O condition and decreased in the S-R condition yielded significant effects throughout the right putamen and globus pallidus (CST, 33, −7, −5) and in the right IPL (CST, 63, −31, 37). The reverse interaction contrast, assessing neural activity that decreased in the R-O condition and increased in the S-R condition, did not reveal any significant effects. Specific assessments of increases vs. decreases revealed decreases across training blocks in the S-R condition, but not in the (exclusively masked) R-O condition, throughout the right putamen and globus pallidus (Fig. 4A). In contrast, significant increases in activity across blocks in the R-O condition, but not in the (exclusively masked) S-R condition, emerged in the right IPL (Fig. 4B).

Neural correlates of devaluation insensitivity

To relate the neuroimaging data to our behavioral effects, we tested whether neural discrimination between instrumental conditions correlated with differences between conditions in the degree of devaluation insensitivity. This was indeed the case. Participants with stronger activation of the tail of the caudate nucleus/thalamus and of the cerebellum, extending into the lingual gyrus, in the S-R relative to the R-O condition during the first two blocks of instrumental learning responded on a greater proportion of devalued trials in the

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

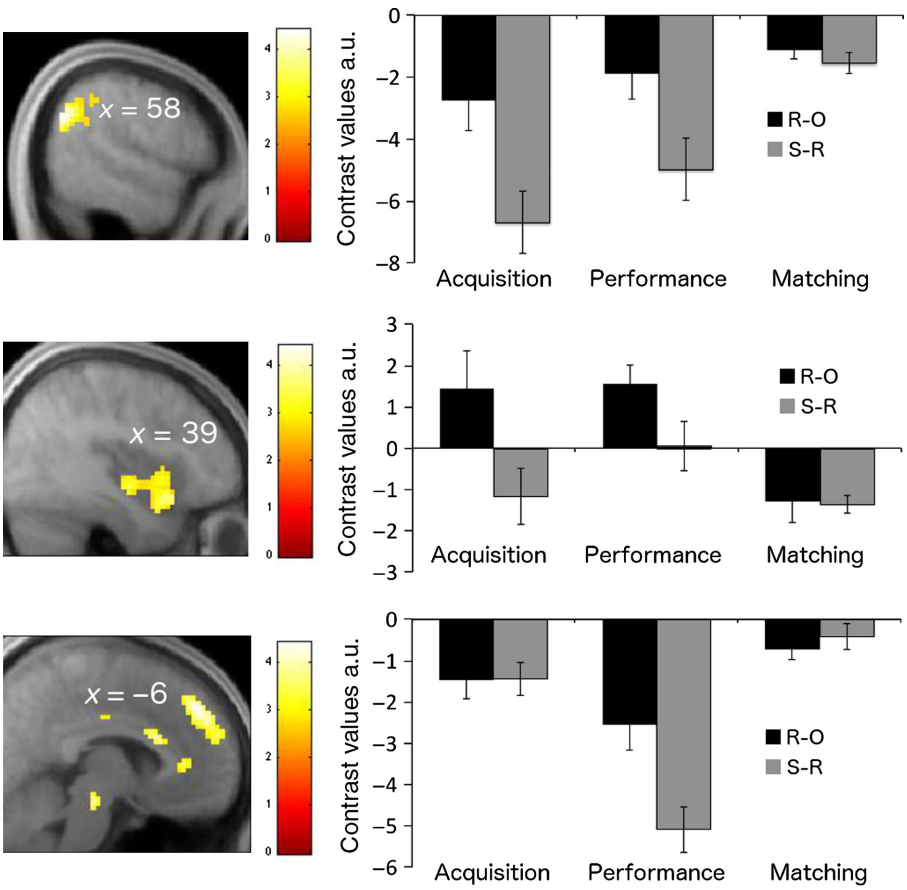


FIG. 3. Areas selectively involved in the R-O condition. Statistical maps show main effects of the R-O > S-R contrast in the posterior ventral IPL (top) and insula (middle), as well as an interaction effect {i.e. [R-O > S-R (performance > acquisition)]} in the dorsomedial prefrontal cortex (bottom). Bar plots show contrast values estimated at leave-one-subject-out coordinates for each instrumental condition, during both acquisition and performance, and for matching controls. Error bars = SEM. a.u., arbitrary units.

S-R condition relative to the R-O condition during the subsequent test phase (Fig. 5A). These effects all survived exclusive masking by an identical contrast applied, with a threshold of 0.1, to the imaging data from the test phase, suggesting that, in these areas, discriminatory activity during early acquisition, but not during test performance, predicted differences in devaluation insensitivity.

Conversely, during the test phase, differences in neural activity between the S-R and R-O conditions in the subgenual cortex (region of interest) and ventral striatum (VS) were positively correlated with the difference between conditions in devaluation insensitivity (Fig. 5B), i.e. greater subgenual and VS activity in the S-R relative to the R-O condition predicted greater devaluation insensitivity in the S-R than R-O condition. We also found a negative correlation with test performance in the right IPL (SVC), such that lesser IPL activity in the S-R relative to the R-O condition predicted greater devaluation insensitivity in the S-R than R-O condition. Again, the specificity of these results was confirmed using exclusive masking by an identical contrast, at a threshold of 0.1, applied to the imaging data from the training phase.

Discussion

In this study we explored the neural substrates of goal-directed and habitual action selection, with a focus on how the recruitment of relevant brain areas might differ across acquisition and implementation of behavioral control. We scanned human participants with functional magnetic resonance imaging as they learned and performed a novel task designed to encourage either goal-directed encoding of the specific outcomes of instrumental responses (R-O condition), or a habitual mapping of responses to antecedent cues (S-R condition). In a subsequent test phase, participants were more likely to respond for a devalued outcome, indicative of habits, in the S-R condition. We found that neural activity in striatal and cortical areas (i) discriminated between our two instrumental conditions, (ii) predicted individual differences in devaluation insensitivity and (iii) did so differentially across acquisition and test performance.

Our results identified several areas in which the degree of neural discrimination between instrumental conditions predicted differences in devaluation insensitivity. Specifically, S-R selective activity in the tail of the caudate nucleus and cerebellum during learning, but not during test performance, predicted greater devaluation insensitivity in the S-R, relative to the R-O, condition. The cerebellum has been strongly implicated in response automatization (Doyon *et al.*, 1998, 2002; Lang & Bastian, 2002; Balsters & Ramnani, 2011), a resistance to dual-task interference postulated to be closely related to, and sometimes treated as synonymous with, habitual performance. Indeed, in the rodent literature, the cerebellum has been directly implicated in habit formation, such that cerebellar lesions abolish devaluation insensitivity in overtrained rats (Callu *et al.*, 2007). However, to our knowledge, no previous study has directly linked the cerebellum to outcome devaluation insensitivity in humans. As with the cerebellum, the tail of the caudate nucleus has been

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

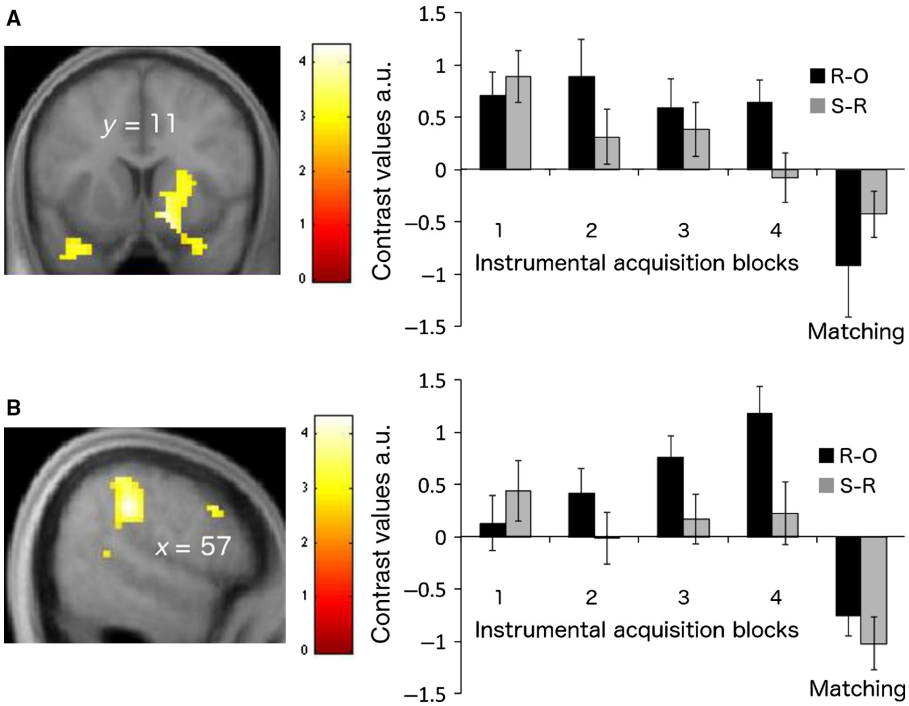


FIG. 4. Discrimination between conditions across blocks of acquisition. (A) Statistical map showing results from disjunction tests assessing decreases across blocks of acquisition in the S-R but not the R-O condition with effects emerging in the right putamen/globus pallidus. (B) Results from a disjunction test of increases across blocks in the R-O but not the S-R condition yielded effects in the IPL. Bar plots show contrast values estimated at leave-one-subject-out coordinates for each instrumental condition in each training block, as well as for matching control conditions. Error bars = SEM. a.u., arbitrary units.

implicated in skill learning (Poldrack & Gabrieli, 2001; Yamamoto *et al.*, 2013). Moreover, single neuron recordings in the monkey tail of the caudate nucleus have revealed that, relative to the body and head of the caudate, this area is specifically involved in encoding stable, non-flexible, values of visual cues (Kim & Hikosaka, 2013; Yamamoto *et al.*, 2013; Hikosaka *et al.*, 2014; Kim *et al.*, 2014), consistent with its role in devaluation insensitivity demonstrated here and elsewhere (Valentin *et al.*, 2007; Liljeholm *et al.*, 2012).

In the subgenual cingulate (Brodmann area 25), S-R selective activity during test performance, but not during learning, correlated with behavioral devaluation insensitivity. Based on its cytoarchitectonic subdivisions and connections, this area has been identified as homologous to the rodent infralimbic cortex. The rodent infralimbic and human subgenual areas both project heavily to the shell of the nucleus accumbens, and both are agranular, relatively poorly laminated areas located ventrally on the medial wall (Gabbott *et al.*, 1997; Ongur & Price, 2000; Ongur *et al.*, 2003). In rodents, the infralimbic cortex has been shown to make specialized contributions to executive control processes facilitating the deployment of habits (Coutureau & Killcross, 2003; Haddon & Killcross, 2011; Smith *et al.*, 2012; Smith & Graybiel, 2013). Consistent with such findings, our results suggest that a putative human homolog of this area does indeed play selective roles in the expression of habits. It should be noted, however, that the rodent infralimbic cortex is primarily known for its involvement in the extinction of Pavlovian responses (Milad & Quirk, 2002; Rhodes & Killcross, 2004, 2007; Santini *et al.*, 2008), whereas the human subgenual cortex has predominantly featured in studies on depression (Greicius *et al.*, 2007; Drevets *et al.*, 2008; Johansen-Berg *et al.*, 2008; Matthews *et al.*, 2009; Keedwell *et al.*, 2010). Future work is needed to reconcile these apparently divergent functions and their potential homology across species.

As with the subgenual cortex, activity in the VS during the test phase, but not during acquisition, predicted devaluation insensitivity. The VS has been frequently shown to support both the acquisition and expression of Pavlovian (stimulus–outcome) associations (Day *et al.*, 2007; Blaiss & Janak, 2009), and to mediate the general motivational influence of such associations on instrumental performance, a phenomenon referred to as Pavlovian-instrumental transfer (Corbit & Balleine, 2011). Of particular importance for interpreting the current findings is the fact that, in rodents, Pavlovian-instrumental transfer appears to selectively influence habitual, rather than goal-directed, responding, i.e. the greater the degree of insensitivity to outcome devaluation, the greater the general motivational influence of Pavlovian cues on instrumental performance (Holland, 2004; Balleine & Ostlund, 2007). We interpret the currently observed activity in the VS as reflecting a greater influence of Pavlovian cues on instrumental responding in the S-R than the R-O condition, and conjecture that this selective engagement of Pavlovian processes supported our behavioral effect.

Contrary to our predictions, we did not find a correlation between activity in the dorsal putamen during early learning and subsequent devaluation insensitivity. We did, however, find S-R selective decreases in right putamen activity across blocks of training, an effect that is consistent with a substantial literature demonstrating a decreased dependence on the putamen with extended (Ungerleider *et al.*, 2002; Poldrack *et al.*, 2005; Turner *et al.*, 2005; Ashby *et al.*, 2010), as well as intermediate (Brovelli *et al.*, 2011), levels of training. Taken together, these results suggest that the contributions of the putamen to automatic and habitual behavioral control may take place early in acquisition, with long-term storage and mediation of well-trained performance occurring elsewhere (Orban *et al.*, 2010). However, some neuroimaging studies have found increases in putamen activity with extended training (Floyer-Lea & Matthews, 2004;

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

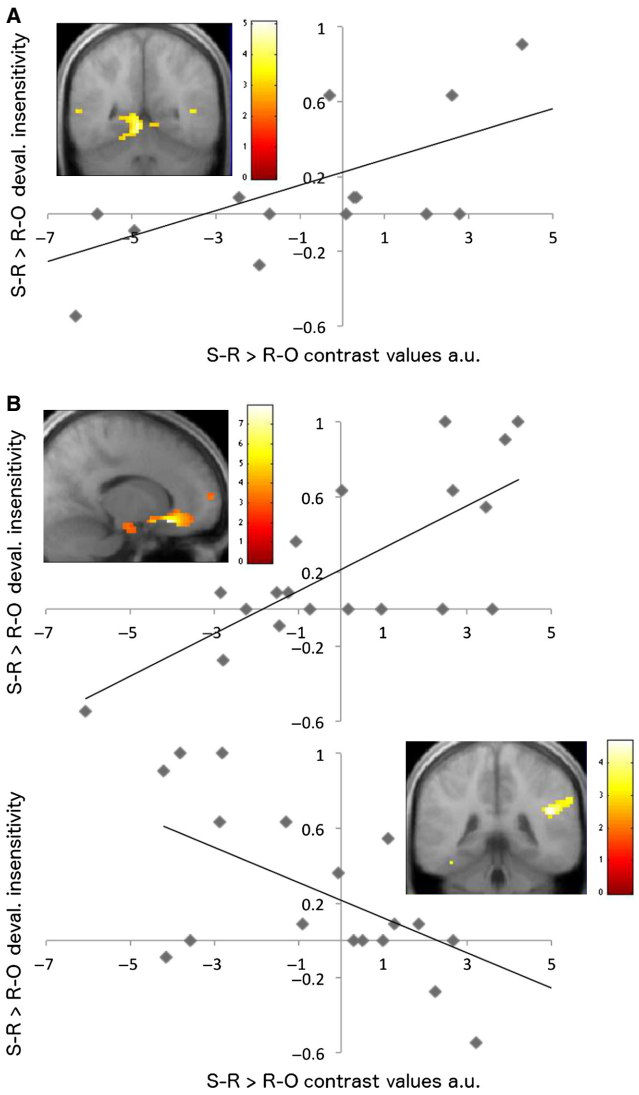


FIG. 5. Correlations between contrast values for the [S-R > R-O (instrumental > control)] contrast and behavioral differences between the S-R and R-O conditions in the proportion of devaluation insensitive responses. Contrast values are extracted from 8 mm spheres centered on the peak voxel in each area. (A) Correlation between contrast values from the first two blocks of instrumental learning (x-axis) and differences between instrumental conditions in the proportion of devaluation insensitive responses (y-axis), showing effects in the cerebellum. (B) Correlation between contrast values from the test phase (x-axis) and concurrent differences between instrumental conditions in the proportion of devaluation insensitive responses (y-axis), showing a positive correlation in the subgenual cortex (top) and a negative correlation in the right IPL (bottom).

Tricomi *et al.*, 2009; Wunderlich *et al.*, 2012). For instance, Tricomi *et al.* (2009) found an increase in activity with overtraining in a far posterior region of the right putamen, concomitant with the development of habitual performance. One key difference between the study of Tricomi *et al.* (2009) and the present one is that the S-R condition in the present study is optimized to generate the rapid acquisition and expression of habitual behavior, whereas in the study of Tricomi *et al.* (2009), behavioral expression of habits (and perhaps also acquisition) emerged much more slowly, becoming evident in behavior only after several days of training. Thus, if the involvement of the putamen in S-R learning dissipates after a period of stable habitual performance, Tricomi *et al.* (2009) may not have sampled behavior beyond that stable period, whereas our accelerated habitual learning paradigm allowed us to do so. It is also worth noting, that, whereas Tricomi *et al.* (2009) reported effects in a small area in the very far posterior putamen, our effects extend throughout the right putamen and globus pallidus.

One feature of the present results is that areas in which discriminatory neural activity was correlated with between-subject variation in devaluation insensitivity did not also show differential main effects in a comparison of S-R and R-O conditions. This may be due in part to the fact that our efforts to differentially encourage S-R and R-O learning resulted only in relative differences in devaluation insensitivity, with the majority of participants showing some degree of sensitivity even in the S-R condition. Thus, neural processes directly related to the degree of habitual responding might not be discernable in analyses that categorically compare conditions. However, not all aspects of encoding S-R associations necessarily contribute to the dominance of such associations over performance. For example, the S-R selective decreases in neural activity across blocks of acquisition, found throughout the right putamen and globus pallidus, may reflect a gradual disengagement of processes that provide critical support during the early development of S-R associations (such as the maintenance of relevant representations, the encoding of discrepancies between attempted and accurate mappings, or the retrieval of response alternatives) but that are supplanted during subsequent performance by processes mediating behavioral control.

We also found selective recruitment by the S-R condition of a lateral middle occipital area that overlaps closely with what has been termed the 'EBA' (Downing *et al.*, 2001). As the name implies, this region is known for its responses to images of human body parts, and has also been frequently implicated in action observation as well as execution (Astafiev *et al.*, 2004; Kuhn *et al.*, 2011; Liljeholm *et al.*, 2012). Notably, studies assessing the role of the EBA in action execution have employed visually guided actions, such that an arbitrary visual stimulus indicates either the location towards which a movement should be directed (Astafiev *et al.*, 2004), or which one of alternative actions should be performed (Kuhn *et al.*, 2011). It is possible therefore that this area is specifically involved in limb movements that are largely stimulus-driven. In the current study, the S-R condition was designed to encourage a mapping of responses to arbitrary antecedent visual cues by decorrelating actions from sensory-specific outcome features (and indeed from any visual features that were intrinsically related to the goal of maintaining beaker fluids). Thus, one possible explanation for the selective recruitment of EBA by the S-R condition is that this area mediates the detection and mapping of arbitrary stimuli to behavioral responses.

The VMPFC has been implicated in goal-directed performance in numerous human neuroimaging studies, and is a proposed homolog of the rodent prelimbic cortex (Balleine & O'Doherty, 2010), lesions of which disrupt the acquisition, but not the expression, of goal-directed performance (Ostlund & Balleine, 2005). Although we did find selective recruitment of the VMPFC by the R-O condition during learning, this effect did not survive the subtraction of matching control conditions. It is possible, of course, that the process supported by the VMPFC, for example assigning values to subgoals, was elicited by the relevant stimuli in the matching task as well as the instrumental task. Another possibility, given our robust effect of experimental phase in the VMPFC, such that activity was greater during early acquisition than during test in both conditions, is that the goal-directed function supported by the VMPFC was present in both instrumental conditions during early acquisition. Further work is needed to arbitrate between these possibilities.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

Further dissociating the contributions of medial prefrontal areas, we found that activity in a more dorsal medial prefrontal region was greater in the R-O than the S-R condition during test performance, but not during early acquisition. This region, along with the adjacent dorsal anterior cingulate, has been implicated in tasks in which the attainment of goals and rewards requires high levels of cognitive control, such as when monitoring for unfavorable outcomes, and during response conflict and decision uncertainty (Ridderinkhof *et al.*, 2004a,b). The currently observed pattern of activity in this area may reflect a decrease in performance monitoring and outcome evaluation during habitual relative to goal-directed control.

In our novel task the stimulus materials were identical across S-R and R-O conditions, but the instrumental contingencies encouraged participants to attend to different visual features (cues vs. beakers). Although we used a simple match-to-sample task to rule out visual processes involved in selectively attending to such features as a source of any imaging effects, there may be aspects of visual attention that are intrinsically related to instrumental responding. For example, we found greater activity in the R-O than the S-R condition in a posterior ventral region of the IPL during both the learning and test phase; during test, discriminatory activity in this area predicted greater sensitivity to devaluation. The posterior ventral region of the IPL has been proposed to function as a 'circuit break' that redirects attention towards behaviorally relevant information that is either exogenously presented (Corbetta & Shulman, 2002; Corbetta *et al.*, 2008; Cabeza *et al.*, 2012) or retrieved into working memory (Cabeza *et al.*, 2012). Importantly, substantial evidence also indicates that the posterior ventral region of the IPL is deactivated when the new information is not relevant to the current task (Shulman *et al.*, 2003). As can be seen in Fig. 3 (top), activity in the posterior ventral region of the IPL appears to be deactivated in both instrumental conditions relevant to the matching control, but markedly more so in the S-R condition. A possible explanation for this pattern of results is that greater suppression was required in the S-R condition because the sensory features of the beaker subgoal, although ultimately irrelevant for response selection, was intimately related to the trial outcome. Indeed, this augmented suppression of sensory-specific outcome features may be a general property of S-R learning, particularly as the retrieval of such features into working memory is considered central to goal-directed encoding (Balleine & Ostlund, 2007).

Consistent with evidence from the rodent literature that goal-directed performance persists in tasks in which alternative actions produce distinct rewards (Colwill & Rescorla, 1985; Holland, 2004), we found that human action selection was more goal-directed in a condition in which instrumental actions obtained unique subgoals. Formally, the degree to which alternative actions yield distinct outcome states can be quantified as the divergence of their outcome probability distributions. In a previous study (Liljeholm *et al.*, 2013), we found that activity in the right anterior IPL (the anterior dorsal supramarginal gyrus) increased with increasing outcome divergence. Likewise, in the current study, training-dependent increases in this area were found in the R-O, but not the S-R, condition, potentially reflecting the incremental increase in divergence as actions became associated with their respective outcome states. The selective recruitment of the IPL by the R-O condition, and the correlation between such discriminatory neural activity and behavioral sensitivity to outcome devaluation, is also consistent with a large body of research implicating this area in various goal-directed processes, including the computation of instrumental contingencies (Seo *et al.*, 2009; Liljeholm *et al.*, 2011), attribution of intent (den Ouden *et al.*, 2005), awareness of agency (Farrer *et al.*, 2008), and mediation of executive function (Friedman & Goldman-Rakic, 1994). Further determination of the exact contributions of the IPL to goal-directed action selection, its role in corollary attentional processes, and in representations of outcome divergence, is an important avenue for future work.

In summary, our results are novel in several critical ways. To our knowledge, we are the first to dissociate the roles of the tail of the caudate nucleus and VS, across learning and test performance, in behavioral insensitivity to outcome devaluation. We are also the first to demonstrate that activity in the IPL, an area that has been previously implicated in several processes closely linked to goal-directed action selection, including the attribution of intent and awareness of agency, predicts sensitivity to outcome devaluation. Finally, we reveal a potential functional homology between the human subgenual and rodent infralimbic cortex in the implementation of habitual control. Taken together, our findings suggest a broad systems division, at the cortical and subcortical levels, between brain areas mediating the acquisition and expression of action–outcome and S-R associations. Notably, a fundamental issue in the search for treatments of behavioral disorders is how to both facilitate the automatization of actions that lead to healthful consequences and abolish well-established deleterious habits. An improved understanding of how distinct neural substrates in the human brain mediate the acquisition vs. expression of habitual control, the aim of the present study, is therefore of significant clinical interest.

Acknowledgements

This work was supported by a grant from the Wellcome Trust and by an NIH grant (DA033077-01) to J.P.O'D. There are no conflicts of interest.

Abbreviations

CST, cluster-size thresholding; EBA, extrastriate body area; IPL, inferior parietal lobule; R-O, Response–Outcome; S-R, Stimulus–Response; SVC, small volume correction; VMPFC, ventromedial prefrontal cortex; VS, ventral striatum.

References

Adams, C.D. (1982) Variations in the sensitivity of instrumental responding to refinforcer devaluation. *Q. J. Exp. Psychol.*, **34**, 77–98.

Adams, C.D. & Dickinson, A. (1981) Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol.*, **33**, 109–121.

Ashby, F.G., Turner, B.O. & Horvitz, J.C. (2010) Cortical and basal ganglia contributions to habit learning and automaticity. *Trends Cogn. Sci.*, **14**, 208–215.

Astafiev, S.V., Stanley, C.M., Shulman, G.L. & Corbetta, M. (2004) Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nat. Neurosci.*, **7**, 542–548.

Balleine, B.W. & Dickinson, A. (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, **37**, 407–419.

Balleine, B.W. & O'Doherty, J.P. (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, **35**, 48–69.

Balleine, B.W. & Ostlund, S.B. (2007) Still at the choice-point: action selection and initiation in instrumental conditioning. *Ann. NY Acad. Sci.*, **1104**, 147–171.

Balsters, J.H. & Ramnani, N. (2011) Cerebellar plasticity and the automation of first-order rules. *J. Neurosci.*, **31**, 2305–2312.

Blaiss, C.A. & Janak, P.H. (2009) The nucleus accumbens core and shell are critical for the expression, but not the consolidation, of Pavlovian conditioned approach. *Behav. Brain Res.*, **200**, 22–32.

Brovelli, A., Nazarian, B., Meunier, M. & Boussaoud, D. (2011) Differential roles of caudate nucleus and putamen during instrumental learning. *NeuroImage*, **57**, 1580–1590.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

Cabeza, R., Ciaramelli, E. & Moscovitch, M. (2012) Cognitive contributions of the ventral parietal cortex: an integrative theoretical account. *Trends Cogn. Sci.*, **16**, 338–352.
Callu, D., Puget, S., Faure, A., Guegan, M. & El Massioui, N. (2007) Habit learning dissociation in rats with lesions to the vermis and the interpositus of the cerebellum. *Neurobiol. Dis.*, **27**, 228–237.
Carelli, R.M. & West, M.O. (1991) Representation of the body by single neurons in the dorsolateral striatum of the awake, unrestrained rat. *J. Comp. Neurol.*, **309**, 231–249.
Carelli, R.M., Wolske, M. & West, M.O. (1997) Loss of lever press-related firing of rat striatal forelimb neurons after repeated sessions in a lever pressing task. *J. Neurosci.*, **17**, 1804–1814.
Colwill, R.M. & Rescorla, R.A. (1985) Instrumental responding remains insensitive to reinfrocer devaluation after extensive training. *J. Exp. Psychol. Anim. B.*, **11**, 520–536.
Corbetta, M. & Shulman, G.L. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.*, **3**, 201–215.
Corbetta, M., Patel, G. & Shulman, G.L. (2008) The reorienting system of the human brain: from environment to theory of mind. *Neuron*, **58**, 306–324.
Corbit, L.H. & Balleine, B.W. (2011) The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. *J. Neurosci.*, **31**, 11786–11794.
Coutureau, E. & Killcross, S. (2003) Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behav. Brain Res.*, **146**, 167–174.
Daw, N.D., Niv, Y. & Dayan, P. (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.*, **8**, 1704–1711.
Day, J.J., Roitman, M.F., Wightman, R.M. & Carelli, R.M. (2007) Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.*, **10**, 1020–1028.
Dickinson, A., Nicholas, D.J. & Adams, C.D. (1983) The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Q. J. Exp. Psychol.-B.*, **35**, 35–51.
Downing, P.E., Jiang, Y., Shuman, M. & Kanwisher, N. (2001) A cortical area selective for visual processing of the human body. *Science*, **293**, 2470–2473.
Doya, K., Samejima, K., Katagiri, K. & Kawato, M. (2002) Multiple model-based reinforcement learning. *Neural Comput.*, **14**, 1347–1369.
Doyon, J., Laforce, R. Jr., Bouchard, G., Gaudreau, D., Roy, J., Poirier, M., Bedard, P.J., Bedard, F. & Bouchard, J.P. (1998) Role of the striatum, cerebellum and frontal lobes in the automatization of a repeated visuomotor sequence of movements. *Neuropsychologia*, **36**, 625–641.
Doyon, J., Song, A.W., Karni, A., Lalonde, F., Adams, M.M. & Ungerleider, L.G. (2002) Experience-dependent changes in cerebellar contributions to motor sequence learning. *Proc. Natl. Acad. Sci. USA*, **99**, 1017–1022.
Draganski, B., Kherif, F., Kloppel, S., Cook, P.A., Alexander, D.C., Parker, G.J., Deichmann, R., Ashburner, J. & Frackowiak, R.S. (2008) Evidence for segregated and integrative connectivity patterns in the human Basal Ganglia. *J. Neurosci.*, **28**, 7143–7152.
Drevets, W.C., Savitz, J. & Trimble, M. (2008) The subgenual anterior cingulate cortex in mood disorders. *CNS Spectr.*, **13**, 663–681.
Esterman, M., Tamber-Rosenau, B.J., Chiu, Y.C. & Yantis, S. (2010) Avoiding non-independence in fMRI data analysis: leave one subject out. *NeuroImage*, **50**, 572–576.
Farrer, C., Frey, S.H., Van Horn, J.D., Tunik, E., Turk, D., Inati, S. & Grafton, S.T. (2008) The angular gyrus computes action awareness representations. *Cereb. Cortex*, **18**, 254–261.
Floyer-Lea, A. & Matthews, P.M. (2004) Changing brain networks for visuomotor control with increased movement automaticity. *J. Neurophysiol.*, **92**, 2405–2412.
Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A. & Noll, D.C. (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.*, **33**, 636–647.
Friedman, H.R. & Goldman-Rakic, P.S. (1994) Coactivation of prefrontal cortex and inferior parietal cortex in working memory tasks revealed by 2DG functional mapping in the rhesus monkey. *J. Neurosci.*, **14**, 2775–2788.
Gabbott, P.L., Dickie, B.G., Vaid, R.R., Headlam, A.J. & Bacon, S.J. (1997) Local-circuit neurones in the medial prefrontal cortex (areas 25, 32 and 24b) in the rat: morphology and quantitative distribution. *J. Comp. Neurol.*, **377**, 465–499.
Glascher, J. (2009) Visualization of group inference data in functional neuroimaging. *Neuroinformatics*, **7**, 73–82.
Greicius, M.D., Flores, B.H., Menon, V., Glover, G.H., Solvason, H.B., Kenna, H., Reiss, A.L. & Schatzberg, A.F. (2007) Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biol. Psychiat.*, **62**, 429–437.
Haddon, J.E. & Killcross, S. (2011) Inactivation of the infralimbic prefrontal cortex in rats reduces the influence of inappropriate habitual responding in a response-conflict task. *Neuroscience*, **199**, 205–212.
Hampton, A.N., Bossaerts, P. & O'Doherty, J.P. (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.*, **26**, 8360–8367.
Hikosaka, O., Kim, H.F., Yasuda, M. & Yamamoto, S. (2014) Basal ganglia circuits for reward value-guided behavior. *Annu. Rev. Neurosci.*, **37**, 289–306.
Holland, P.C. (2004) Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *J. Exp. Psychol. Anim. B.*, **30**, 104–117.
Jog, M.S., Kubota, Y., Connolly, C.I., Hillegaart, V. & Graybiel, A.M. (1999) Building neural representations of habits. *Science*, **286**, 1745–1749.
Johansen-Berg, H., Gutman, D.A., Behrens, T.E., Matthews, P.M., Rushworth, M.F., Katz, E., Lozano, A.M. & Mayberg, H.S. (2008) Anatomical connectivity of the subgenual cingulate region targeted with deep brain stimulation for treatment-resistant depression. *Cereb. Cortex*, **18**, 1374–1383.
Keedwell, P.A., Drapier, D., Surguladze, S., Giampietro, V., Brammer, M. & Phillips, M. (2010) Subgenual cingulate and visual cortex responses to sad faces predict clinical outcome during antidepressant treatment for depression. *J. Affect. Disorders*, **120**, 120–125.
Killcross, S. & Coutureau, E. (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex*, **13**, 400–408.
Kim, H.F. & Hikosaka, O. (2013) Distinct basal ganglia circuits controlling behaviors guided by flexible and stable values. *Neuron*, **79**, 1001–1010.
Kim, H.F., Ghazizadeh, A. & Hikosaka, O. (2014) Separate groups of dopamine neurons innervate caudate head and tail encoding flexible and stable value memories. *Front. Neuroanatomy*, **8**, 120.
Kuhn, S., Keizer, A., Rombouts, S.A. & Hommel, B. (2011) The functional and neural mechanism of action preparation: roles of EBA and FFA in voluntary action control. *J. Cognitive Neurosci.*, **23**, 214–220.
Lang, C.E. & Bastian, A.J. (2002) Cerebellar damage impairs automaticity of a recently practiced movement. *J. Neurophysiol.*, **87**, 1336–1347.
Lee, S.W., Shimojo, S. & O'Doherty, J.P. (2014) Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, **81**, 687–699.
Liljeholm, M., Tricomi, E., O'Doherty, J.P. & Balleine, B.W. (2011) Neural correlates of instrumental contingency learning: differential effects of action-reward conjunction and disjunction. *J. Neurosci.*, **31**, 2474–2480.
Liljeholm, M., Molloy, C.J. & O'Doherty, J.P. (2012) Dissociable brain systems mediate vicarious learning of stimulus-response and action-outcome contingencies. *J. Neurosci.*, **32**, 9878–9886.
Liljeholm, M., Wang, S., Zhang, J. & O'Doherty, J.P. (2013) Neural correlates of the divergence of instrumental probability distributions. *J. Neurosci.*, **33**, 12519–12527.
Maldjian, J.A., Laurienti, P.J., Kraft, R.A. & Burdette, J.H. (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, **19**, 1233–1239.
Matthews, S., Simmons, A., Strigo, I., Gianaros, P., Yang, T. & Paulus, M. (2009) Inhibition-related activity in subgenual cingulate is associated with symptom severity in major depression. *Psychiat. Res.*, **172**, 1–6.
Milad, M.R. & Quirk, G.J. (2002) Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature*, **420**, 70–74.
O'Doherty, J.P., Deichmann, R., Critchley, H.D. & Dolan, R.J. (2002) Neural responses during anticipation of a primary taste reward. *Neuron*, **33**, 815–826.
Ongur, D. & Price, J.L. (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex*, **10**, 206–219.
Ongur, D., Ferry, A.T. & Price, J.L. (2003) Architectonic subdivision of the human orbital and medial prefrontal cortex. *J. Comp. Neurol.*, **460**, 425–449.
Orban, P., Peigneux, P., Lungu, O., Albouy, G., Breton, E., Laberenne, F., Benali, H., Maquet, P. & Doyon, J. (2010) The multifaceted nature of the relationship between performance and brain activity in motor sequence learning. *NeuroImage*, **49**, 694–702.
Ostlund, S.B. & Balleine, B.W. (2005) Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. *J. Neurosci.*, **25**, 7763–7770.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

den Ouden, H.M., Frith, U. & Frith, C.S.J.B. (2005) Thinking about intentions. *NeuroImage*, **28**, 787–796.
Poldrack, R.A. & Gabrieli, J.D. (2001) Characterizing the neural mechanisms of skill learning and repetition priming: evidence from mirror reading. *Brain*, **124**, 67–82.
Poldrack, R.A., Sabb, F.W., Foerde, K., Tom, S.M., Asarnow, R.F., Bookheimer, S.Y. & Knowlton, B.J. (2005) The neural correlates of motor skill automaticity. *J. Neurosci.*, **25**, 5356–5364.
Rhodes, S.E. & Killcross, S. (2004) Lesions of rat infralimbic cortex enhance recovery and reinstatement of an appetitive Pavlovian response. *Learn. Memory*, **11**, 611–616.
Rhodes, S.E. & Killcross, A.S. (2007) Lesions of rat infralimbic cortex enhance renewal of extinguished appetitive Pavlovian responding. *Eur. J. Neurosci.*, **25**, 2498–2503.
Ridderinkhof, K.R., van den Wildenberg, W.P., Segalowitz, S.J. & Carter, C.S. (2004a) Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain Cognition*, **56**, 129–140.
Ridderinkhof, K.R., Ullsperger, M., Crone, E.A. & Nieuwenhuis, S. (2004b) The role of the medial frontal cortex in cognitive control. *Science*, **306**, 443–447.
Santini, E., Quirk, G.J. & Porter, J.T. (2008) Fear conditioning and extinction differentially modify the intrinsic excitability of infralimbic neurons. *J. Neurosci.*, **28**, 4028–4036.
Seo, H., Barraclough, D.J. & Lee, D. (2009) Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. *J. Neurosci.*, **29**, 7278–7289.
Shulman, G.L., McAvoy, M.P., Cowan, M.C., Astafiev, S.V., Tansy, A.P., d'Avossa, G. & Corbetta, M. (2003) Quantitative analysis of attention and detection signals during visual search. *J. Neurophysiol.*, **90**, 3384–3397.
Smith, K.S. & Graybiel, A.M. (2013) A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron*, **79**, 361–374.
Smith, K.S., Virkud, A., Deisseroth, K. & Graybiel, A.M. (2012) Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex. *Proc. Natl. Acad. Sci. USA*, **109**, 18932–18937.
Tanaka, S.C., Balleine, B.W. & O'Doherty, J.P. (2008) Calculating consequences: brain systems that encode the causal effects of actions. *J. Neurosci.*, **28**, 6750–6755.
Tang, C., Pawlak, A.P., Prokopenko, V. & West, M.O. (2007) Changes in activity of the striatum during formation of a motor habit. *Eur. J. Neurosci.*, **25**, 1212–1227.
Tricomi, E., Balleine, B.W. & O'Doherty, J.P. (2009) A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.*, **29**, 2225–2232.
Turner, R.S., McCairn, K., Simmons, D. & Bar-Gad, I. (2005) Sequential motor behavior and the basal ganglia. In Bolam, J.P., Ingham, C.A. & Magill, P.J. (Eds), *The Basal Ganglia VIII (Advances in Behavioral Biology)*. Springer, New York, pp. 563–574.
Ungerleider, L.G., Doyon, J. & Karni, A. (2002) Imaging brain plasticity during motor skill learning. *Neurobiol. Learn. Mem.*, **78**, 553–564.
Valentin, V.V., Dickinson, A. & O'Doherty, J.P. (2007) Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.*, **27**, 4019–4026.
Ward, B.D. (2000) Simultaneous inference for fMRI data [Computer software manual]. AFNI 3dDeconvolve Documentation, Medical College of Wisconsin. [Internet] Available http://afni.nimh.nih.gov/afni/doc/manual/AlphaSim.
de Wit, S., Watson, P., Harsay, H.A., Cohen, M.X., van de Vijver, I. & Ridderinkhof, K.R. (2012) Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *J. Neurosci.*, **32**, 12066–12075.
Wunderlich, K., Dayan, P. & Dolan, R.J. (2012) Mapping value based planning and extensively trained choice in the human brain. *Nat. Neurosci.*, **15**, 786–791.
Yamamoto, S., Kim, H.F. & Hikosaka, O. (2013) Reward value-contingent changes of visual responses in the primate caudate tail associated with a visuomotor skill. *J. Neurosci.*, **33**, 11227–11238.
Yin, H.H., Knowlton, B.J. & Balleine, B.W. (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.*, **19**, 181–189.
Yin, H.H., Knowlton, B.J. & Balleine, B.W. (2005) Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *Eur. J. Neurosci.*, **22**, 505–512.

© 2015 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

Copyright of European Journal of Neuroscience is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

http://dx.doi.org/10.5665/sleep.3836

Hippocampal Substructural Vulnerability to Sleep Disturbance and Cognitive Impairment in Patients with Chronic Primary Insomnia: Magnetic Resonance Imaging Morphometry

Eun Yeon Joo, MD, PhD[1]; Hosung Kim, PhD[2]; Sooyeon Suh, PhD[3,4]; Seung Bong Hong, MD, PhD[1]

[1]Sleep Center, Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Samsung Biomedical Research Institute, Seoul, Korea; [2]Department of Neurology and Brain Imaging Center, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada; [3]Korea University Ansan Hospital, Human Genome Institute, Korea University, Seoul, Korea; [4]Stanford University, Department of Psychiatry, Redwood City, CA

Study Objectives: Despite compelling evidence from animal studies indicating hippocampal subfield-specific vulnerability to poor sleep quality and related cognitive impairment, there have been no human magnetic resonance imaging (MRI) studies investigating the relationship between hippocampal subfield volume and sleep disturbance. Our aim was to investigate the pattern of volume changes across hippocampal subfields in patients with primary insomnia relative to controls.

Design: Pointwise morphometry allowed for volume measurements of hippocampal regions on T1-weighted MRI.

Setting: University hospital.

Patients: Twenty-seven unmedicated patients (age: 51.2 ± 9.6 y) and 30 good sleepers as controls (50.4 ± 7.1 y).

Interventions: N/A.

Measurements: We compared hippocampal subfield volumes between patients and controls and correlated volume with clinical and neuropsychological features in patients.

Results: Patients exhibited bilateral atrophy across all hippocampal subfields ($P < 0.05$ corrected). Cornu ammonis (CA) 1 subfield atrophy was associated with worse sleep quality (higher Pittsburgh Sleep Quality Index and higher arousal index of polysomnography) ($r < -0.45$, $P < 0.005$). The volume of the combined region, including the dentate gyrus (DG) and CA3-4, negatively correlated with verbal memory, verbal information processing, and verbal fluency in patients ($|r| > 0.45$, $P < 0.05$). Hemispheric volume asymmetry of this region (left smaller than right) was associated with impaired verbal domain functions ($r = 0.50$, $P < 0.005$).

Conclusion: Hippocampal subfield atrophy in chronic insomnia suggests reduced neurogenesis in the dentate gyrus (DG) and neuronal loss in the cornu ammonis (CA) subfields in conditions of sleep fragmentation and related chronic stress condition. Atrophy in the CA3-4-DG region was associated with impaired cognitive functions in patients. These observations may provide insight into pathophysiological mechanisms that make patients with chronic sleep disturbance vulnerable to cognitive impairment.

Keywords: chronic insomnia, hippocampus, memory, morphometry, sleep, surface analysis

Citation: Joo EY, Kim H, Suh S, Hong SB. Hippocampal substructural vulnerability to sleep disturbance and cognitive impairment in patients with chronic primary insomnia: magnetic resonance imaging morphometry. *SLEEP* 2014;37(7):1189-1198.

INTRODUCTION

Chronic primary insomnia, also termed psychophysiological insomnia, is a disorder characterized by chronically disturbed sleep and sleep loss, nonrefreshing sleep, and heightened arousal in bed; these symptoms have a profound effect on psychological well-being, as well as on social and cognitive functioning.[1] Patients with chronic primary insomnia commonly have daytime cognitive impairments and deficits in memory consolidation during sleep compared to good sleepers.[2,3] Negative association of disturbed sleep with performance on memory tasks[2,4,5] implicates hippocampal dysfunction in patients. The hippocampus consists of cytoarchitectonically distinctive and functionally specialized subfields,[6–8] the cornu ammonis sectors (CA1-4), the dentate gyrus (DG), and the subiculum. The hippocampus is a target for glucocorticoids and is thereby vulnerable to stress-induced responses that affect both the structure and function of the hippocampus.[9] Acute or prolonged psychosocial stress is associated with structural changes of the CA1 or CA3 neurons[10,11] as well as the DG in mammals.[12,13] Individual hippocampal subfields are responsible for distinct cognitive functions, and differentially vulnerable to neurological and neuropsychiatric conditions.[14–18] Experimental sleep deprivation and fragmentation affect neurogenesis and cell proliferation.[19,20] These studies have suggested that chronic psychological stress conditions driven by sleep fragmentation in patients with chronic insomnia may contribute to alterations in the flow of information through the hippocampal network because of an anomaly in the synaptic pathways, which is reflected by subfield-specific cytoarchitectonic damage, ultimately yielding a negative influence on cognition.[21]

Sophisticated neuroimaging techniques allow for the *in vivo* visualization of human brain anatomy with exquisite detail and the quantification of morphological changes. Recent chronic insomnia studies have analyzed hippocampal volume (HV) on magnetic resonance imaging (MRI),[22–24] but findings have been inconsistent. Although a first study observed bilaterally reduced

Submitted for publication October, 2013
Submitted in final revised form January, 2014
Accepted for publication February, 2014

Address correspondence to: Seung Bong Hong, MD, PhD, Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; E-mail: sbhong@skku.edu, smcsbhong@naver.com

Downloaded from https://academic.oup.com/sleep/article-abstract/37/7/1189/2709372
by guest
on 01 February 2018

HV in patients compared to good sleepers,[23] two recent studies did not replicate this finding.[22,24] Moreover, our group[22] found that memory impairments in patients with chronic insomnia are related to a decrease in HV. This discrepancy may be related not only to the low sensitivity of global hippocampal volumetry, but also the variability of the hippocampal segmentation between studies.[25] So far, there has been only one study that examined the regional specificity of hippocampal changes in patients and good sleepers, which revealed no significant difference between groups.[26] The automated subfield volumetry used in the study may not have been sufficiently sensitive to visualize hippocampal subfields at a low resolution MRI (1 mm^3). Alternatively, vertex (= point)-wise morphometry based on a surface extracted from the manual segmentation of the whole hippocampus[27–29] has been a surrogate to manual subfield volumetry, which can only be done on high-field (> 3 Tesla) and high-resolution MRI.[30–33] These surface-based approaches, which generate deformation maps that encode individual local shape variation relative to the template, have successfully identified hippocampal subfield pathology in various brain disorders.[27,28,34–36]

The objective of the current study is to investigate the pattern of volume changes across hippocampal subfields in patients compared to controls. Furthermore, we assessed the relationships between subfield volumes and clinical and neuropsychological parameters in patients. To this end, we quantified local volume changes using our previously developed surface-based framework. This procedure extracts the hippocampal boundary from manual labeling using spherical mapping to guarantee intersubject point correspondence.[37] Furthermore, it computes surface-based, point-wise jacobian determinants that represent local volume changes of a given hippocampus relative to the mean shape of all healthy hippocampi.[38] Compared to conventional surface shape analyses,[23,33] our technique, derived from other approaches,[37,39] allows the measurement of actual changes in volume independent of positional changes.[38,40]

METHODS

Forty patients with complaints of sleep onset and/or maintenance insomnia were recruited from a sleep disorder clinic in a university hospital setting (Samsung Seoul Hospital) located in Seoul, South Korea. Patients included in the current study must have met the following criteria: age 40–70 y, meet diagnostic criteria for chronic primary insomnia according to the International Classification of Sleep Disorders-2,[41] and have an insomnia duration of more than 1 y.

Good sleeper controls were recruited through an advertisement in the local community. Their age, sex, and education status did not differ from the patients with chronic insomnia ($P > 0.2$). All participants underwent a comprehensive clinical interview, during which a history of medical and sleep disorders was acquired. History of psychiatric disorders was assessed using the Structured Clinical Interview for Diagnostic and Statistical Manual for Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR).[42] All participants completed a comprehensive physical and neurological examination, overnight polysomnography (PSG), neuropsychological tests, a sleep diary for at least 2 w, and a number of questionnaires assessing sleep quality and mood status.

Exclusion criteria were: (1) controls with an average nightly sleep time of less than 7 h over the recent 2 w based on sleep diaries; (2) obstructive sleep apnea (OSA, apnea-hypopnea index greater than 5/h); 3) moderate to severe periodic limb movement during sleep (PLMS, total PLM index > 25/h); (4) circadian rhythm sleep disorder determined by sleep-wake cycles of sleep diaries; (5) hypertension, diabetes, heart disease, or respiratory diseases; (6) history of cerebrovascular disease; (7) diagnosis of other neurological (neurodegenerative diseases, epilepsy, head injury) or psychiatric diseases (psychosis, current depression); (8) alcohol or illicit drug abuse or current intake of psychoactive medications; and (9) a structural lesion on brain MRI. To avoid the possible effect of medications on the results, four patients who had a history of taking antidepressants, anxiolytics, or sleeping pills were excluded. Four patients were excluded because of definite sleep disorders confirmed by PSG; three with moderate to severe OSA (apnea-hypopnea index, 17.5–40.2/h) and one with PLMS (total index 50.1/h, movement arousal index 10.5/h). Five patients who showed diffuse brain atrophy, white matter changes, or lacunar infarctions on brain MRIs were also excluded. Twenty-seven patients and 30 controls were included in the study. Of these subjects, 17 insomnia patients and 16 good sleepers were also included in a previous study.[22] Informed consent was obtained from all participants, and the institutional review board of Samsung Seoul Hospital authorized the study protocol and design.

Overnight Polysomnography

The day before the sleep study, subjects were asked to refrain from drinking alcohol or caffeinated beverages. Sleep studies were recorded using a Remlogic (Embla Systems, Denver, CO, USA). We have previously provided a detailed description of the test procedures.[39]

Neuropsychological Assessment

Subjects underwent a battery (2.5 h) of neuropsychological tests and an individual standardized intelligence test. Neuropsychological tests assessed six broad domains: working memory (digit span test from the Wechsler Memory Scale-Revised and the Corsi Block tapping tests, forward and backward); executive functioning (Trail-Making Tests A and B and the Stroop test), verbal information processing (digit symbol test), verbal memory (Korean California Verbal Test), visual memory (Rey Complex Figure Test) and verbal fluency (controlled oral word association test). Information on the protocol of the neuropsychological assessments can be found elsewhere.[22]

Self-Report Questionnaires

The Pittsburgh Sleep Quality Index (PSQI) measures the quality and disturbances of sleep during the past month.[44] A global score greater than 5 indicates poor sleep quality.

The Insomnia Severity Index (ISI) assesses subjective symptoms of insomnia, including the degree of distress caused by this particular sleep complaint.[45] A higher score represents greater insomnia severity. Scores above 14 are generally consistent with clinical levels of insomnia.[46]

The Beck Depression Inventory-II (BDI-II) measures the severity of depressive symptoms.[47] Higher scores reflect greater levels of depressive symptoms.

Downloaded from https://academic.oup.com/sleep/article-abstract/37/7/1189/2709372
by guest
on 01 February 2018

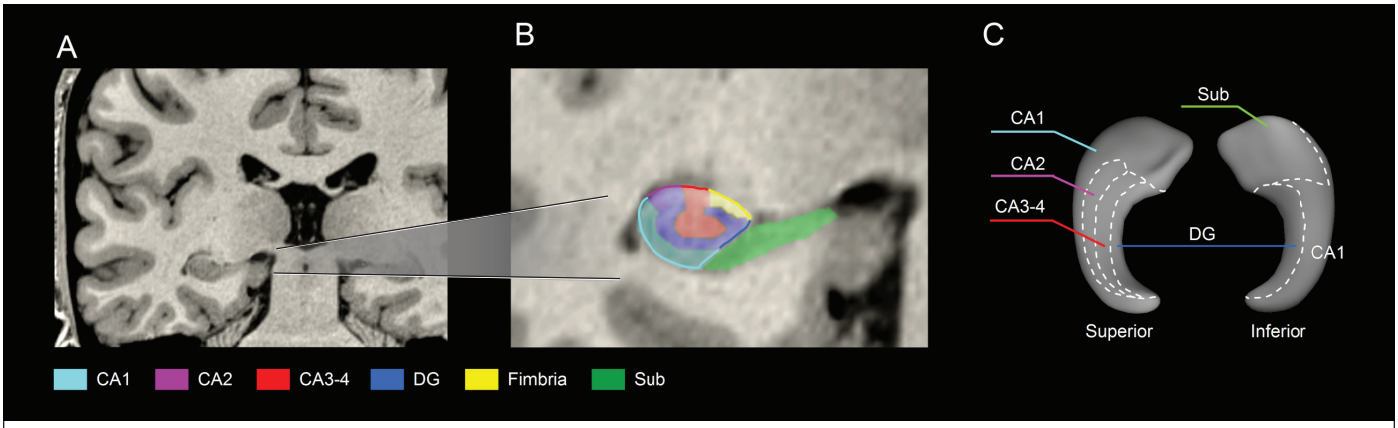


Figure 1—Surface-based subfield atlas. **(A, B)** Individual MRI on which hippocampal subfields are manually and schematically delineated based on Duvernoy atlas.[6] Our segmentation protocol excludes the subiculum at the level of hippocampal body and tail.[20] **(C)** Surface of the entire hippocampus for the same individual. By registering the given individual surface to the template representing healthy population, we obtain the borders between subfields (white dot lines) on the surface. CA, cornu ammonis; DG, dentage gyrus; Sub, subiculum

MRI Acquisition and Image Processing

MRI scanning was performed with a GE Signa 1.5-Tesla scanner (GE Medical Systems, Milwaukee, WI, USA). T1-weighted coronal spoiled gradient recalled (SPGR) MRI was performed with the following scanning variables: 1.6-mm thickness, no gap, 124 slices, repetition time (TR)/echo time (TE) = 30/7 ms, fractional anisotropy (FA) = 45°, number of excitation (NEX) = 1, matrix size = 256 × 192, and field of view = 22 × 22 cm. The voxel dimensions in the SPGR MRI images were 0.86 × 0.86 × 1.6 mm. Each image underwent automated correction for intensity nonuniformity and intensity standardization.[48]

To control for differences in brain volume, magnetic resonance images were registered into the the Montreal Neurological Institute/International Consortium of Brain Mapping standard brain template (MNI/ICBM) 152 nonlinear template[45] using a nine-parameter linear transformation.[50]

Labels of the hippocampus were obtained by manual segmentation on T1-weighted MRI according to a previous protocol[51] and used in an insomnia study,[23] which includes the DG and CA and excludes the subiculum (Figures 1A, 1B). We tested intrarater reliability on five randomly chosen control subjects. The agreement based on kappa statistics was excellent ($k = 0.94 \pm 0.02$), demonstrating that this protocol is highly reproducible.

Surface-Based Mapping of Mesiotemporal Atrophy

We used a previously validated surface-based approach[38] to measure local volume changes. This approach computes jacobian determinants on surface-based displacement vectors between a given subject and a template surface.[37] Briefly, manually segmented labels were first minimally smoothed and then converted to surface meshes for which a spherical parameterization (SPHARM) was computed using area-preserving, distortion-minimizing mapping (Figures 2A-2C). Based on a uniform icosahedron-subdivision of the SPHARM, we obtained a point distribution model (PDM), allowing shape-inherent point (1,002 vertices) correspondences across subjects (Figures 2D, 2E). Each individual SPHARM-PDM surfaces were rigidly aligned to a template constructed from the mean surface of controls and patients with respect to the centroid and the longitudinal axis of the first order ellipsoid (Figure 2E).[52] Vertexwise displacement vectors were then calculated between each individual and the template (Figure 2F).[37] Displacement vectors were diffused within the volume enclosed by the surface using a heat equation, yielding a displacement vector field. To assess local volume changes, we finally calculated jacobian determinants from these vector fields.[38] By projecting these jacobian determinants back onto the surface using trilinear interpolation and subtracting one from them, we obtained a metric of growth ($J > 0$) or shrinkage ($J < 0$) in a unit-size cube defined on each vertex (Figure 2F). We have previously shown that volume change at a vertex is driven by the structure underneath the surface.[38] Given that the surface-based analysis was performed in stereotaxic space, correction for differences in overall intracranial volume was unnecessary.

Statistical Analysis

To compare demographics, self-reported questionnaire scores, PSG data, and neuropsychological parameters between patients and controls, one-way analysis of variance was used. Six composite scores were computed from the neuropsychological tests by transforming all scores into standardized Z scores and then averaging them to represent each domain: working memory, executive functioning, verbal information processing, verbal memory, visual memory, and language processing.

Analyses to determine HV was performed using SurfStat toolbox[53] for Matlab (R2007a, The Mathworks, Natick, MA, USA). Global HV and surface-based local volumes were standardized relative to the distribution of controls using a Z-score transformation. Because no laterality of volume changes was expected in relation to insomnia,[23] we averaged the left and right volumes. In particular, local volumes were averaged at the same location between the left and right hemispheres. All the following statistical tests included BDI-II as a covariate because patients showed elevated depressive mood although this was clinically insignificant. (1) Group comparison: We compared differences in global volume between patients and controls using two-tailed *t*-tests. To assess local volume differences, we repeated the same test on vertex-wise jacobian determinants at each vertex. To assess a possible hemispheric

Downloaded from https://academic.oup.com/sleep/article-abstract/37/7/1189/2709372
by guest
on 01 February 2018

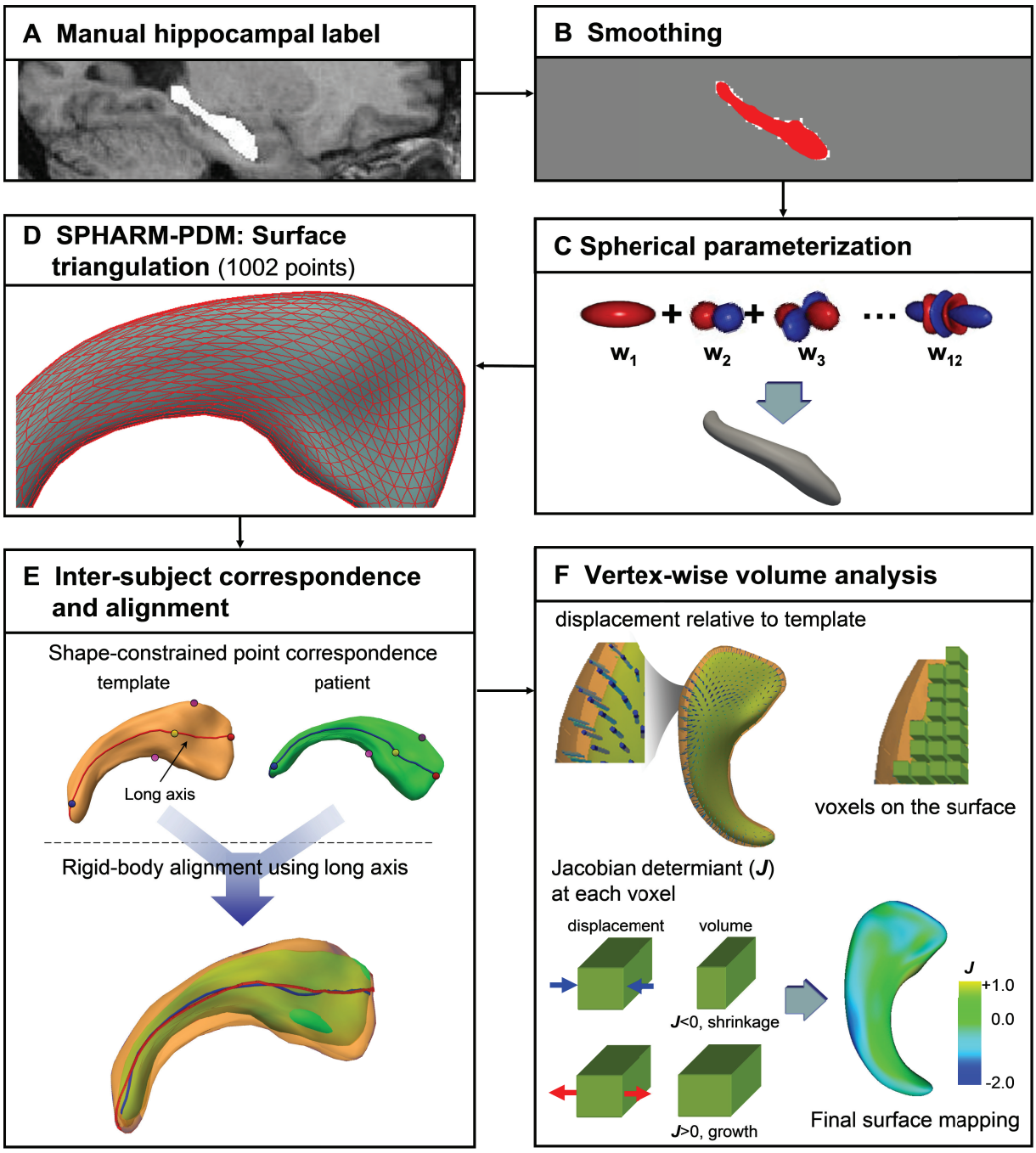


Figure 2—Flowchart of the surface-based analysis procedure. Manually segmented labels **(A)** are first minimally smoothed **(B)** and then converted to surface meshes for which a spherical parameterization (SPHARM) is computed using area-preserving, distortion-minimizing mapping **(C)**. Based on a uniform icosahedron-triangulation of the SPHARM, we obtain a point distribution model (PDM) **(D)**, allowing shape-inherent point (= vertices) correspondences across subjects **(E)** (top). Individual SPHARM-PDM surfaces are rigidly aligned to the template shape **(E)** (bottom). Vertex-wise displacement vectors are then calculated between each individual and the template **(F)** (top). Displacement vectors are diffused to voxels **(F)** (top) enclosed by the surface using a heat equation, yielding voxelwise vector field. To assess local volume changes, we calculate jacobian determinants (J) from this vector field. By projecting this metric back onto the surface using tri-linear interpolation, we finally obtain a metric of growth (J > 0) or shrinkage (J < 0) at each vertex **(F)** (bottom).

asymmetry in the changes in patients compared to controls, we first computed asymmetry in volume differences between hemispheres (i.e., 2 × (L − R) / (L + R) where L and R denote the left and right hemisphere) and further performed a *t*-test on this metric between patients and controls. (2) Reliability test: To test the reproducibility of our surface-based analysis, we performed a bootstrap analysis in which we iteratively selected random subsamples of the whole dataset (i.e., 10,000 times) and performed group comparisons at each iteration. To evaluate the consistency of patterns of volume change, we mapped the probability of having significant volume changes across all iterations on the template surface. (3) Association with clinical and neuropsychological parameters: To assess the association of volume changes with clinical demographics, PSG

Downloaded from https://academic.oup.com/sleep/article-abstract/37/7/1189/2709372
by guest
on 01 February 2018

and neuropsychological parameters, linear models were applied to compute Pearson correlation coefficient *r* while controlling for age, sex, and depressive mood. For local volume changes, the correlation was performed on individual measurements averaged within each significant cluster, as determined by group comparison. (4) Correction for multiple comparisons: In all vertex-wise surface analyses, significances were thresholded using the false discovery rate (FDR) procedure,[54] with FDR < 0.05. In the volumetric analysis, the Bonferroni adjustment was applied to control the family-wise error rate. (5) Localization of findings: We schematically outlined hippocampal subfields (Figure 1) on the surface template based on Duvernoy atlas[6] and its surface reconstructions that were previously published.[27,55] The segmentation protocol excluded the subiculum as the subiculum has been shown to be least affected by sleep deprivation in animal studies[19,20] and an MRI study of patients with posttraumatic stress disorder.[56] Furthermore, the one insomnia study that found significantly smaller HV among patients also excluded the subiculum.[23] We include the subiculum at the level of the hippocampal head because it was difficult to distinguish it from the hippocampus proper at our current spatial resolution. An approximate border was defined between the CA1 and DG on the inferomedial surface of the template. Although the hippocampal head contains all subfields, the portion of DG is relatively small and the DG and the CA2-4 situate in the middle.[6,57] Thus, we kept only the CA1 subfield on our surface atlas in the hippocampal head as in the previous atlases.[27,55]

Table 1—Characteristics of patients with chronic primary insomnia and controls

	Patients (n = 27)	Controls (n = 30)	P
Demographics			
Mean age, y	51.2 ± 9.6	50.4 ± 7.1	0.7
Female, n (%)	25 (93)	28 (93)	0.7
Duration of insomnia, y	8.4 ± 9.1	–	–
Body mass index, kg/m^2	25.5 ± 2.5	26.3 ± 1.3	0.1
Education, y	13.6 ± 1.9	14.4 ± 1.9	0.1
Beck Depression Inventory-II	10.1 ± 5.2	5.0 ± 4.1	< 0.001
Sleep questionnaires			
Pittsburgh Sleep Quality Index	14.9 ± 4.6	3.2 ± 0.9	< 0.001
Insomnia Severity Index	19.0 ± 5.2	2.4 ± 1.5	0.001
Overnight polysomnography			
Total sleep time, min	350.1 ± 60.0	409.4 ± 29.7	< 0.001
Sleep latency, min	25.5 ± 27.7	8.7 ± 6.2	0.002
REM latency, min	140.7 ± 77.3	83.1 ± 28.9	< 0.001
Sleep efficiency, %	80.8 ± 10.0	89.2 ± 5.4	< 0.001
WASO, min	59.2 ± 38.7	41.2 ± 25.3	0.04
Arousal index, /h	13.3 ± 6.6	9.4 ± 6.0	0.03
Apnea-hypopnea index, /h	2.5 ± 2.7	1.1 ± 1.6	0.02
PLMS index, /h	4.0 ± 6.0	1.8 ± 6.3	0.2
Movement Arousal Index, /h	1.0 ± 1.7	2.6 ± 7.6	0.3
N 1, %	13.3 ± 6.0	13.2 ± 6.0	1
N 2, %	62.7 ± 12.3	53.8 ± 6.0	0.001
N3, %	4.9 ± 6.8	9.0 ± 7.4	0.03
REM sleep, %	19.2 ± 9.0	23.6 ± 3.6	0.02

PLMS, periodic limb movement in sleep; REM, rapid eye movement; WASO, wakefulness after sleep onset.

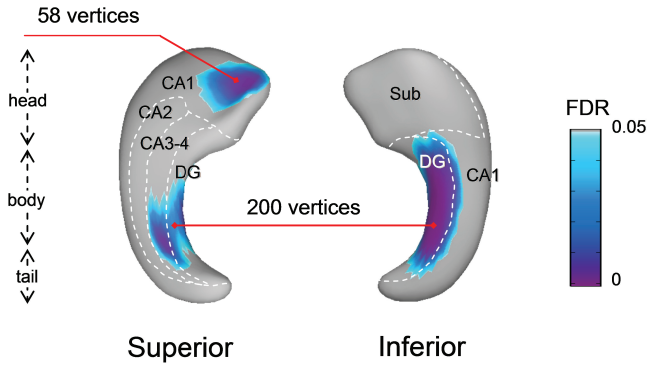


Figure 3—Vertexwise group comparison between patients with chronic primary insomnia and controls. Regions of volume decrease in patients relative to controls are shown. Significances are thresholded at FDR < 0.05.

RESULTS

Demographics, Self-Report Questionnaires, and Overnight Sleep Studies

All subjects were right-handed. There was no difference in the mean age or sex ratio between patients and controls (P > 0.7). Subjective sleep quality was determined by the PSQI, which was assessed during the 2 w prior to the study; as expected, the mean PSQI scores were significantly higher in patients than in the controls. Patients also reported severe symptoms and had significantly higher scores on the depressive mood index compared to controls (P < 0.001), although they reported no current or lifetime psychiatric disorders during the structured clinical interview. Detailed results are summarized in Table 1.

Neuropsychological Tests

Patients had significantly lower scores in working memory, verbal information processing, verbal memory, verbal fluency, and visual memory (P < 0.0001), but not executive functioning (P = 0.6) when compared to controls. The details of neuropsychological tests are summarized in Table 2.

Global and Local Volume Changes in Patients with Chronic Insomnia Relative to Controls

A two-tailed *t*-test uncovered a significant decrease in the HV of patients versus controls (left: 2980 ± 283 versus 3197 ± 337 mm^3; right: 3079 ± 298 versus 3247 ± 404 mm^3, P < 0.05). This change was hemispherically symmetric; asymmetry between patients and controls did not differ (P > 0.2, i.e., patients presented a 3% volume difference between the left and right hippocampi and controls also presented a similar asymmetry (2%). Results of vertex-wise comparisons between groups are shown in Figure 3. In patients, hippocampal atrophy was identified in all subfields. Vertex-wise asymmetry replicated this symmetric change (FDR > 0.1). We separately mapped regional volume changes in patients

Downloaded from https://academic.oup.com/sleep/article-abstract/37/7/1189/2709372
by guest
on 01 February 2018

Table 2—Neuropsychological analyses of patients with chronic primary insomnia and controls

	Patients (n = 27)	Good sleepers (n = 30)	P
Working memory composite score	-0.17 ± 0.42	0.36 ± 0.36	< 0.001
Digit span forward	7.6 ± 2.0	9.5 ± 2.0	
Digit span backward	5.4 ± 1.5	8.6 ± 2.1	
Corsi block forward	7.8 ± 1.9	10.7 ± 1.8	
Corsi block backward	7.5 ± 1.6	9.9 ± 0.9	
Trail-making test A	57.4 ± 52.4	34.8 ± 10.4	
Trail-making test B	125.9 ± 94.2	82.0 ± 30.1	
Executive functioning composite score	-0.05 ± 0.87	0.05 ± 0.49	0.6
Stroop word correct responses	112 ± 1.0	106.0 ± 14.9	
Stroop word correct responses time	112.7 ± 11.8	108.4 ± 12	
Verbal information processing composite score	-0.47 ± 1.03	0.4 ± 0.8	< 0.001
Digit symbol test	49.5 ± 12.5	65.2 ± 12.1	
Verbal memory composite score	-0.53 ± 1.05	0.48 ± 0.64	< 0.001
Korean California Verbal Test			
Total	45.4 ± 12.2	57.2 ± 7.4	
Short delay free recall	9.4 ± 2.8	12.5 ± 3.0	
Long delay free recall	9.7 ± 3.3	12.7 ± 2.7	
Recognition	14.4 ± 1.3	15.1 ± 1.3	
Verbal fluency composite score	-0.22 ± 0.83	0.20 ± 0.71	< 0.0001
Controlled Oral Word Association Test			
Phonetic word fluency	27.9 ± 11.9	32.1 ± 12.7	
Semantic word fluency	31.6 ± 8.2	36.2 ± 7.7	
Animal	15.2 ± 4.81	20.8 ± 5.1	
Supermarket	16.3 ± 5.2	15.4 ± 4.8	
Visual memory composite score	-0.41 ± 0.77	0.37 ± 0.62	< 0.001
Rey Complex Figure Test			
Copy	30.9 ± 6.82	34.6 ± 2.5	
Immediate recall	13.8 ± 7.7	21.6 ± 7.3	
Delayed recall	13.7 ± 7.8	20.8 ± 7.1	
Recognition	19.5 ± 1.9	20.8 ± 1.9	

compared to controls in left and right hippocampi (Figure S1, supplemental material). This confirmed that the pattern of regional atrophy was similar between the left and right hippocampi. The largest cluster of atrophy was detected at the level of the hippocampal body and tail (FDR < 0.005, 200 vertices) and located medially, mainly within the region corresponding to the combined region of CA2-4 and DG (Figure 1A). Atrophy at the level of the head was present on the superomedial surface corresponding to CA1 (FDR < 0.05, 58 vertices). There was no local volume increase in patients relative to controls (FDR > 0.1). Bootstrap reproducibility analysis revealed that the proposed surface-based analytic platform reproduced the same patterns of atrophy in more than 85% of the random subsampling iterations (Figure S2, supplemental material).

Association Between Hippocampal Volume, and Clinical and Neuropsychological Parameters

We found that age, sex, and education did not correlate with HV in patients or controls (P > 0.2). Depressive mood (BDI-II) was associated with smaller HV (r = -0.39, P < 0.05). In patients, hippocampal atrophy were associated with worse sleep quality (PSQI) (r = -0.40, P < 0.05) and a higher arousal index on the overnight PSG (r = -0.44, P < 0.05).

Surface-based analysis revealed that the degree of volume loss in the cluster located at CA1 subfield (Figure 3A) was associated with a higher arousal index (r < -0.50, P < 0.005) as well as higher PSQI (r = -0.45, P < 0.01, Figure 3B). No significant correlations were found between surface-based volume and insomnia duration or other PSG parameters (|r| < 0.23, P > 0.2).

The atrophy of DG and CA2-4 (Figure 4A) was associated with decreased verbal memory, processing, and fluency scores (r > 0.51, P < 0.005; Figure 4C). Volume asymmetry (left smaller than right) of this area was also associated with impaired verbal processing (r = 0.50, P < 0.005).

The CA1 atrophy of the hippocampal head (Figure 4A) did not correlate with any neuropsychological parameters (r < 0.30; P > 0.1).

DISCUSSION

The current study investigated hippocampal subfield volume changes in patients with chronic primary insomnia compared to controls, and the association of specific hippocampal subfield volume differences with clinical and neuropsychological parameters.

Reduced Hippocampal Volume in Patients with Chronic Primary Insomnia

Our study uncovered bilateral atrophy across subfields in patients. Surface-based analysis revealed that this atrophy mainly corresponded to the region CA2-4-DG of the hippocampal body and tail, and CA1 subfield of the hippocampal head.

Animal studies have reported that sustained sleep fragmentation reduced neurogenesis in hippocampal subfields, particularly in the DG.[19,20] In patients with chronic insomnia, sleep fragmentation correlates with higher cortisol levels in the evening and nighttime periods.[58] The hippocampus contains a higher density of glucocorticoid receptors, which make it more vulnerable to long-term stress than most other brain areas.[59] Chronic elevations in corticosterone have been shown to suppress synaptic plasticity in the DG.[60] Chronic restraint and stress-related steroids have been shown to cause atrophy of apical dendrites of CA3 pyramidal neurons[61,62] and result in the depletion and reorganization of synaptic vesicles in mossy fiber terminals in the CA3.[61] Supporting this notion, it has been reported that patients with Cushing disease[62] and those with posttraumatic stress disorder,[63] both of whom have increased levels of cortisol, have hippocampal atrophy.

Patients in our study suffered from long-term sleep disturbances (mean 8.4 y) and their PSG findings supported evidence

Downloaded from https://academic.oup.com/sleep/article-abstract/37/7/1189/2709372
by guest
on 01 February 2018

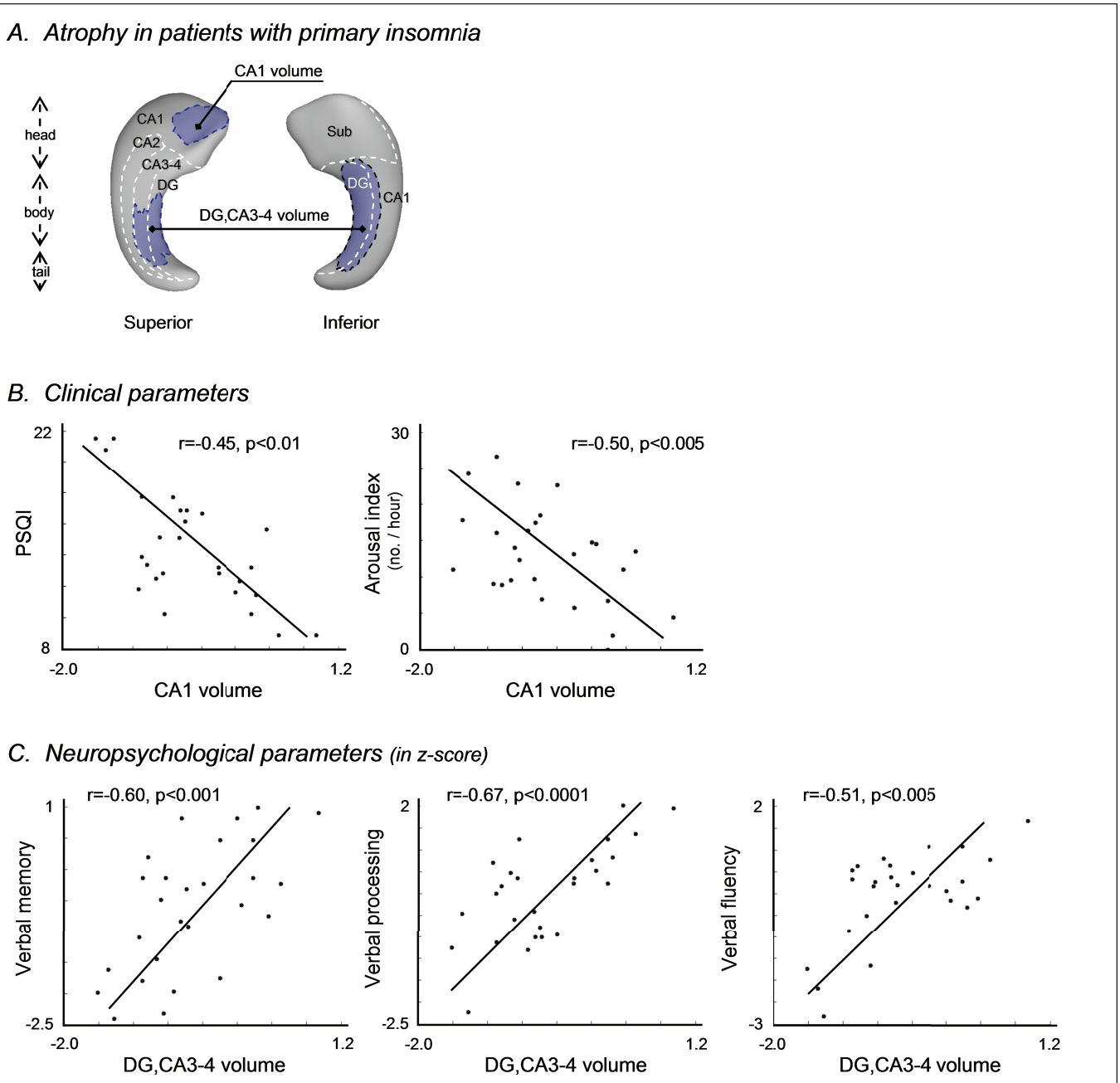


Figure 4—Association between hippocampal subfield volume and clinical **(B)** and neuropsychological **(C)** parameters in patients with chronic primary insomnia. For each cluster representing significant volume loss in patients relative to controls **(A)**, its mean volume is correlated with given clinical or neuropsychological parameter while controlling age, sex and depressive mood. Linear regression models are plotted for significant correlations. PSQI, Pittsburgh Sleep Quality Index.

of their sleep fragmentation, indexed by increased wakefulness after sleep onset, frequent arousals, and reduced sleep efficiency compared to controls. Higher arousal indices on PSG and PSQI scores were significantly associated with smaller CA1 subfield volumes in patients. It has been found that melatonin production is significantly diminished in chronic insomnia patients with sleep fragmentation.[23] Furthermore, melatonin has demonstrated protective effects in CA1 neuronal density in animal models.[64]

Our findings therefore suggest that a combined effect of disturbed sleep in patients and related chronic stress conditions may be related to the structural alteration of subfields of the hippocampal formation.

Relationship Between Hippocampal Atrophy and Cognitive Function in Patients

In the current study, patients showed cognitive impairment in verbal information processing, verbal memory, verbal processing, and visual memory compared to controls. These parameters correlated significantly with volume alterations in the cluster overlapping with the CA2-4-DG region.

The DG has been of particular interest in human cognition because dentate granule cells that are generated continuously in the adult mammalian brain[65,66] play an important role in regulating cognition[67] and storing new memories.[68] Sleep disturbance suppresses adult DG neurogenesis, especially with

Downloaded from https://academic.oup.com/sleep/article-abstract/37/7/1189/2709372
by guest
on 01 February 2018

prolonged sleep disruption, which may have implications for patients with chronic insomnia who have had long-term sleep disturbance.[19,20,69,70] An MRI study that performed volumetry in patients with posttraumatic stress disorder also showed a negative correlation between insomnia symptom severity and volumes of the CA3/DG subfields.[56] It is therefore tempting to speculate that structural alterations of specific hippocampal subfields, which are secondary to prolonged sleep disruption, may lead to the impairment of cognition in patients.

The relationship between hippocampal atrophy and cognitive impairment may also be explained by a disorganization of the hippocampal circuit. The hippocampal formation is organized as a mainly unidirectional circuit made up of multiple subregions — the entorhinal cortex, the DG, the CA1 and CA3 subfields, and the subiculum.[18] These subregions differ in their cellular organization and connectivity[71] and thus, potentially make distinct contributions to the known stages of memory processing.[62,72] A previous neuroimaging study has also demonstrated that early subregions in the hippocampal circuit (DG and CA2-3) are more selectively active during episodic memory creation, whereas the next subfield in the circuit (subiculum) is more involved in the retrieval process.[73] Atrophy found in our study may be macroscopic, which is indicative of cytoarchitectonic damage in the CA2-4-DG and therefore may impair the encoding process rather than the retrieval process. This hypothesis is consistent with patients with chronic insomnia complaining of difficulties with episodic memory.[4] An MRI volumetric study of patients with subjective memory impairment confirmed that performance in memory acquisition through training is related to the volume of a combined region of CA2-4 and DG in the hippocampus.[74]

Our analysis revealed that the volume atrophy of the CA3-4-DG subfield was particularly associated with verbal-specific domains (information processing, memory, and fluency). Notably, volume asymmetry (left smaller than right) of this region in patients was also associated with impaired verbal processing. Sleep loss impedes the spontaneous generation of words during verbal word fluency tasks and the articulation of speech during vocalized reading tasks in healthy subjects[75] and patients with chronic insomnia.[76] The hippocampus is a functional network hub for memory. Information is continuously exchanged in a network of brain areas centered at the hippocampus and including the neocortex and other key structures such as the amygdala and the striatum.[77] Verbal memory is processed asymmetrically between the cerebral cortices. Encoding of verbal memory preferentially relies on left hemispheric brain regions including the hippocampus.[78] Functional MRI study have confirmed that the left hippocampus is highly activated during verbal memory tasks.[79] Moreover, the hippocampus is integrally involved in the retrieval and binding of information, critical for language processing.[80] The extent of hippocampus remaining after an anterior temporal lobectomy in patients with left mesiotemporal sclerosis was significantly related to verbal fluency after surgery.[81] Therefore, our finding that left hippocampal atrophy is related to poor performance of verbal memory and verbal fluency in patients may be explained by a pathological disruption in the brain network involved in verbal-specific domain processing.

Comparison with Other Hippocampal Volumetry Studies

Manual volumetry has been considered a standard approach to assess hippocampal structural alterations in patients with chronic insomnia. Despite a generally accepted sensitivity of this technique, because of the precise delineation of the anatomy, inconsistent reports have been published. This inconsistency may caused by variations in which substructures were included in the delineation of the hippocampus. Our study and a previous study[23] excluded the subiculum from delineation and found significantly smaller HV among patients, whereas other studies[22,24] included the subiculum and did not uncover these differences. Moreover, we studied a large sample size compared to all previous studies except one.[82] The increased significance of our results may therefore be a consequence of increased statistical power due to our large database as well as the exclusion of the subiculum, which occupies more than 20% of the HV.[32,33] The subiculum is least likely to be affected by sleep deprivation, as suggested in animal studies[19,20] and an MRI study of patients with posttraumatic stress disorder in which low sleep quality was assessed.[56]

To the best of our knowledge, our study is the first to localize HV changes in patients with primary insomnia using a surface-based analytic model. This advanced approach, relying on mesh parameterization and the use of smoothness constraints, is known to enable the localization of volume changes as small as the voxel size.[36,37,83] Indeed, our surface mapping unveils region-specific volume reductions in the hippocampus related to chronic sleep disturbance. A vertex-wise correlation of subfield volumes with clinical and neuropsychological parameters allows us to find relationships among structural changes, insomnia severity, and cognitive impairments. However, our technique has a limitation in that it differs from manual hippocampal subfield volumetry.[30–33] It is difficult to distinguish subfields at the current image resolution (1 mm^3), thus we adopted a pointwise analytic platform and interpreted changes in patients by corresponding them to a schematic atlas of the subfields based on a histology study[6] as well as previously published surface-based atlases.[27,55] The localization of atrophy to subfields in patients was thus approximate.

CONCLUSION

Our study provides evidence of disturbed sleep in patients with primary insomnia affecting the hippocampal substructure. Our primary findings indicate that patients exhibit bilateral atrophy corresponding to the CA2–4-DG of the hippocampal body and tail and CA1 of the hippocampal head. A decrease in sleep quality was closely related to the degree of hippocampal CA1 atrophy. Furthermore, atrophy located in the CA2-4-DG region was associated with impaired cognitive functions in patients with PI. These observations may provide insight into the pathophysiological mechanisms that make patients with chronic sleep disturbances vulnerable to cognitive impairment.

ACKNOWLEDGMENT

The authors are grateful to Jessie Kulaga-Yoskovitz, MSc at McGill University for her proofreading and editorial help. Drs. Joo and Kim contributed equally to the writing of this manuscript.

Downloaded from https://academic.oup.com/sleep/article-abstract/37/7/1189/2709372
by guest
on 01 February 2018

DISCLOSURE STATEMENT

This was not an industry supported study. This study was supported by a grant of the Korean Health Technology R&D Project, Ministry for Health and Welfare, Republic of Korea (No. A110097). The authors have indicated no financial conflicts of interest.

REFERENCES

1. Roth T, Roehrs T, Pies R. Insomnia: pathophysiology and implications for treatment. Sleep Med Rev 2007;11:71–9.
2. Backhaus J, Junghanns K, Born J, Hohaus K, Faasch F, Hohagen F. Impaired declarative memory consolidation during sleep in patients with primary insomnia: Influence of sleep architecture and nocturnal cortisol release. Biol Psychiatry 2006;60:1324–30.
3. Fulda S, Schulz H. Cognitive dysfunction in sleep disorders. Sleep Med Rev 2001;5:423–45.
4. Fortier-Brochu E, Beaulieu-Bonneau S, Ivers H, Morin CM. Insomnia and daytime cognitive performance: a meta-analysis. Sleep Med Rev 2012;16:83–94.
5. Oosterman JM, van Someren EJ, Vogels RL, Van Harten B, Scherder EJ. Fragmentation of the rest-activity rhythm correlates with age-related cognitive deficits. J Sleep Res 2009;18:129–35.
6. Duvernoy HM. The Human Hippocampus: An Atlas of Applied Anatomy,, 3rd ed. New York, NY: Springer-Verlag, 2005.
7. Gilbert PE, Kesner RP, Lee I. Dissociating hippocampal subregions: double dissociation between dentate gyrus and CA1. Hippocampus 2001;11:626–36.
8. Witter MP, Wouterlood FG, Naber PA, Van Haeften T. Anatomical organization of the parahippocampal-hippocampal network. Ann N Y Acad Sci 2000;911:1–2.
9. Leuner B, Shors TJ. Stress, anxiety, and dendritic spines: what are the connections? Neuroscience 2013;251:108–19.
10. Magariños AM, McEwen BS, Flügge G, Fuchs E. Chronic psychosocial stress causes apical dendritic atrophy of hippocampal CA3 pyramidal neurons in subordinate tree shrews. J Neurosci 1996;16:3534–40.
11. McKittrick CR, Magarinos AM, Blanchard DC, Blanchard RJ, McEwen BS, Sakai RR. Chronic social stress reduces dendritic arbors in CA3 of hippocampus and decreases binding to serotonin transporter sites. Synapse 2000;36:85–94.
12. Ferland CL, Hawley WR, Puckett RE, et al. Sirtuin activity in dentate gyrus contributes to chronic stress-induced behavior and extracellular signal-regulated protein kinases 1 and 2 cascade changes in the hippocampus. Biol Psychiatry 2013;74:927–35.
13. Sebastian V, Estil JB, Chen D, Schrott LM, Serrano PA. Acute physiological stress promotes clustering of synaptic markers and alters spine morphology in the hippocampus. PLoS One 2013;8:e79077.
14. Brickman AM, Stern Y, Small SA. Hippocampal subregions differentially associate with standardized memory tests. Hippocampus 2011;21:923–8.
15. Kent K, Hess K, Tonegawa S, Small SA. CA3 NMDA receptors are required for experience-dependent shifts in hippocampal activity. Hippocampus 2007;17:1003–11.
16. Nakazawa K, Quirk MC, Chitwood RA, et al. Requirement for hippocampal CA3 NMDA receptors in associative memory recall. Science 2002;297:211–8.
17. Schobel SA, Lewandowski NM, Corcoran CM, et al. Differential targeting of the CA1 subfield of the hippocampal formation by schizophrenia and related psychotic disorders. Arch Gen Psychiatry 2009;66:938–46.
18. Wu W, Brickman AM, Luchsinger J, et al. The brain in the age of old: the hippocampal formation is targeted differentially by diseases of late life. Ann Neurol 2008;64:698–706.
19. Guzman-Marin R, Bashir T, Suntsova N, Szymusiak R, McGinty D. Hippocampal neurogenesis is reduced by sleep fragmentation in the adult rat. Neuroscience 2007;148:325–33.
20. Tung A, Takase L, Fornal C, Jacobs B. Effects of sleep deprivation and recovery sleep upon cell proliferation in adult rat dentate gyrus. Neuroscience 2005;134:721–3.
21. Killgore WD. Effects of sleep deprivation on cognition. Prog Brain Res 2010;185:105–29.
22. Noh HJ, Joo EY, Kim ST, et al. The relationship between hippocampal volume and cognition in patients with chronic primary insomnia. J Clin Neurol 2012;8:130–8.
23. Riemann D, Voderholzer U, Spiegelhalder K, et al. Chronic insomnia and MRI-measured hippocampal volumes: a pilot study. Sleep 2007;30:955–8.
24. Winkelman JW, Benson KL, Buxton OM, et al. Lack of hippocampal volume differences in primary insomnia and good sleeper controls: an MRI volumetric study at 3 Tesla. Sleep Med 2010;11:576–82.
25. Wang Z, Neylan TC, Mueller SG, et al. Magnetic resonance imaging of hippocampal subfields in posttraumatic stress disorder. Arch Gen Psychiatry 2010;67:296–303.
26. Spiegelhalder K, Regen W, Baglioni C, et al. Insomnia does not appear to be associated with substantial structural brain changes. Sleep 2013;36:731–7.
27. Hogan RE, Wang L, Bertrand ME, et al. MRI-based high-dimensional hippocampal mapping in mesial temporal lobe epilepsy. Brain 2004;127:1731–40.
28. Joseph J, Warton C, Jacobson SW, et al. Three-dimensional surface deformation-based shape analysis of hippocampus and caudate nucleus in children with fetal alcohol spectrum disorders. Hum Brain Mapp 2014;35:659–72.
29. Zierhut KC, Grassmann R, Kaufmann J, Steiner J, Bogerts B, Schiltz K. Hippocampal CA1 deformity is related to symptom severity and antipsychotic dosage in schizophrenia. Brain 2013;136:804–14.
30. Kerchner GA, Hess CP, Hammond-Rosenbluth KE, et al. Hippocampal CA1 apical neuropil atrophy in mild Alzheimer disease visualized with 7-T MRI. Neurology 2010;75:1381–7.
31. Mueller SG, Chao LL, Berman B, Weiner MW. Evidence for functional specialization of hippocampal subfields detected by MR subfield volumetry on high resolution images at 4 T. Neuroimage 2011;56:851–7.
32. Winterburn JL, Pruessner JC, Chavez S, et al. A novel in vivo atlas of human hippocampal subfields using high-resolution 3T magnetic resonance imaging. Neuroimage 2013;74:254–65.
33. Wisse LE, Gerritsen L, Zwanenburg JJ, et al. Subfields of the hippocampal formation at 7 T MRI: in vivo volumetric assessment. Neuroimage 2012;61:1043–9.
34. Narr KL, Thompson PM, Szeszko P, et al. Regional specificity of hippocampal volume reductions in first-episode schizophrenia. Neuroimage 2004;21:1563–75.
35. Nicolson R, DeVito TJ, Vidal CN, et al. Detection and mapping of hippocampal abnormalities in autism. Psychiatry research. Neuroimaging 2006;148:11–21.
36. Styner M, Lieberman JA, Pantazis D, Gerig G. Boundary and medial shape analysis of the hippocampus in schizophrenia. Med Image Anal 2004;8:197–203.
37. Styner M, Oguz I, Xu S, Brechbühler C, Pantazis D, Levitt JJ, Shenton ME, Gerig G. Framework for the statistical shape analysis of brain structures using SPHARM-PDM. Insight J 2006;(1071):242-50.
38. Kim H, Besson P, Colliot O, Bernasconi A, Bernasconi N. Surface-based vector analysis using heat equation interpolation: a new approach to quantify local hippocampal volume changes. Med Image Comput Comput Assist Interv 2008;5241:1008–15.
39. Wang Y, Song Y, Rajagopalan P, et al. Surface-based TBM boosts power to detect disease effects on the brain: an N=804 ADNI study. Neuroimage 2011;56:1993–2010.
40. Kim H, Mansi T, Bernasconi N. Disentangling hippocampal shape anomalies in epilepsy. Front Neurol 2013;4:131.
41. American Academy of Sleep Medicine. International Classification of Sleep Disorders, 2nd ed.:diagnostic and coding manual. Westchester, IL: American Academy of Sleep Medicine, 2005.
42. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Text Revision. Washington, DC: American Psychiatric Association, 2000.
43. Joo EY, Noh HJ, Koo DL, et al. Brain gray matter deficits in patients with chronic primary insomnia. Sleep 2013;36:999–1007.
44. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28:193–213.
45. Bastien CH, Vallieres A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001;2:297–307.
46. Smith MT, Wegener ST. Measures of sleep: The Insomnia Severity Index, Medical Outcomes Study (MOS) Sleep Scale, Pittsburgh Sleep Diary (PSD), and Pittsburgh Sleep Quality Index (PSQI). Arthritis Care Res 2003;49:S184–96.

Downloaded from https://academic.oup.com/sleep/article-abstract/37/7/1189/2709372
by guest
on 01 February 2018

47. Beck AT, Brown G, Steer RA, Eidelson JI, Riskind JH. Differentiating anxiety and depression: a test of the cognitive content-specificity hypothesis. J Abnorm Psychol 1987;96:179–83.
48. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 1998;17:87–97.
49. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL. Unbiased average age-appropriate atlases for pediatric studies. Neuroimage 2011;54:313–27.
50. Collins DL, Neelin P, Peters TM, Evans AC. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 1994;18:192–205.
51. Watson C, Jack CR Jr, Cendes F. Volumetric magnetic resonance imaging. Clinical applications and contributions to the understanding of temporal lobe epilepsy. Arch Neurol 1997;54:1521–31.
52. Gerig G, Styner M, Jones D, Weinberg DR, Lieberman J. Shape analysis of brain ventricles using SPHARM. MMBIA 2001:171–8.
53. Chung MK, Worsley KJ, Nacewicz BM, Dalton KM, Davidson RJ. General multivariate linear modeling of surface shapes using SurfStat. Neuroimage 2010;53:491–505.
54. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc 1995;57:289–300.
55. Sabattoli F, Boccardi M, Galluzzi S, Treves A, Thompson PM, Frisoni GB. Hippocampal shape differences in dementia with Lewy bodies. Neuroimage 2008;41:699–705.
56. Neylan TC, Mueller SG, Wang Z, et al. Insomnia severity is associated with a decreased volume of the CA3/dentate gyrus hippocampal subfield. Biol Psychiatry 2010;68:494–6.
57. Insausti AM, Amaral DG. Hippocampal formation. In: Paxinos G, Mai JK, eds. The Human Nervous System. San Diego, CA: Elsevier, 2004:871–914.
58. Rodenbeck A, Huether G, Ruther E, Hajak G. Interactions between evening and nocturnal cortisol secretion and sleep parameters in patients with severe chronic primary insomnia. Neurosci Lett 2002;324:159–63.
59. Joels M, Krugers H, Karst H. Stress-induced changes in hippocampal function. Prog Brain Res 2008;167:3–15.
60. Watanabe Y, McEwen BS. Effects of glucocorticoids on hippocampal long-term potentiation. Hippocampus 1993;3:183–92.
61. Magarinos AM, Verdugo JM, McEwen BS. Chronic stress alters synaptic terminal structure in hippocampus. Proc Natl Acad Sci U S A 1997;94:14002–8.
62. Campbell S, Macqueen G. The role of the hippocampus in the pathophysiology of major depression. J Psychiatry Neurosci 2004;29:417–26.
63. Karl A, Schaefer M, Malta LS, Dorfel D, Rohleder N, Werner A. A meta-analysis of structural brain abnormalities in PTSD. Neurosci Biobehav Rev 2006;30:1004–31.
64. Cho S, Joh TH, Baik HH, Dibinis C, Volpe BT. Melatonin administration protects CA1 hippocampal neurons after transient forebrain ischemia in rats. Brain Res 1997;755:335–8.
65. Cameron HA, McKay RD. Adult neurogenesis produces a large pool of new granule cells in the dentate gyrus. J Comp Neurol 2001;435:406–17.
66. Ming GL, Song H. Adult neurogenesis in the mammalian brain: significant answers and significant questions. Neuron 2011;70:687–702.
67. Zhao N, Zhong C, Wang Y, et al. Impaired hippocampal neurogenesis is involved in cognitive dysfunction induced by thiamine deficiency at early pre-pathological lesion stage. Neurobiol Dis 2008;29:176–85.
68. Kitamura T, Saitoh Y, Takashima N, et al. Adult neurogenesis modulates the hippocampus-dependent period of associative fear memory. Cell 2009;139:814–27.
69. Hairston IS, Little MT, Scanlon MD, et al. Sleep restriction suppresses neurogenesis induced by hippocampus-dependent learning. J Neurophysiol 2005;94:4224–33.
70. Mueller AD, Pollock MS, Lieblich SE, Epp JR, Galea LA, Mistlberger RE. Sleep deprivation can inhibit adult hippocampal neurogenesis independent of adrenal stress hormones. Am J Physiol Regul Integr Comp Physiol 2008;294:R1693–1703.
71. Amaral DG. Emerging principles of intrinsic hippocampal organization. Curr Opin Neurobiol 1993;3:225–9.
72. Eichenbaum H. A cortical-hippocampal system for declarative memory. Nat Rev Neurosci 2000;1:41–50.
73. Eldridge LL, Engel SA, Zeineh MM, Bookheimer SY, Knowlton BJ. A dissociation of encoding and retrieval processes in the human hippocampus. J Neurosci 2005;25:3280–6.
74. Engvig A, Fjell AM, Westlye LT, Skaane NV, Sundseth O, Walhovd KB. Hippocampal subfield volumes correlate with memory training benefit in subjective memory impairment. Neuroimage 2012;61:188–94.
75. Harrison Y, Horne JA. Sleep deprivation affects speech. Sleep 1997;20:871–7.
76. Fernandez-Mendoza J, Calhoun S, Bixler EO, et al. Insomnia with objective short sleep duration is associated with deficits in neuropsychological performance: a general population study. Sleep 2010;33:459–65.
77. Battaglia FP, Benchenane K, Sirota A, Pennartz CM, Wiener SI. The hippocampus: hub of brain network communication for memory. Trends Cogn Sci 2011;15:310–8.
78. Poldrack RA, Gabrieli JD. Memory and the brain: what's right and what's left? Cell 1998;93:1091–3.
79. Jansen A, Sehlmeyer C, Pfleiderer B, et al. Assessment of verbal memory by fMRI: lateralization and functional neuroanatomy. Clin Neurol Neurosurg 2009;111:57–62.
80. Duff MC, Brown-Schmidt S. The hippocampus and the flexible use and processing of language. Front Hum Neurosci 20125;6:69.
81. Bonelli SB, Thompson PJ, Yogarajah M, et al. Imaging language networks before and after anterior temporal lobe resection: results of a longitudinal fMRI study. Epilepsia 2012;53:639–50.
82. Spiegelhalder K, Regen W, Baglioni C, et al. Insomnia does not appear to be associated with substantial structural brain changes. Sleep 2013;36:731–7.
83. Qiu A, Younes L, Miller MI, Csernansky JG. Parallel transport in diffeomorphisms distinguishes the time-dependent pattern of hippocampal surface deformation due to healthy aging and the dementia of the Alzheimer's type. Neuroimage 2008;40:68–76.

Downloaded from https://academic.oup.com/sleep/article-abstract/37/7/1189/2709372
by guest
on 01 February 2018

SUPPLEMENTAL MATERIAL

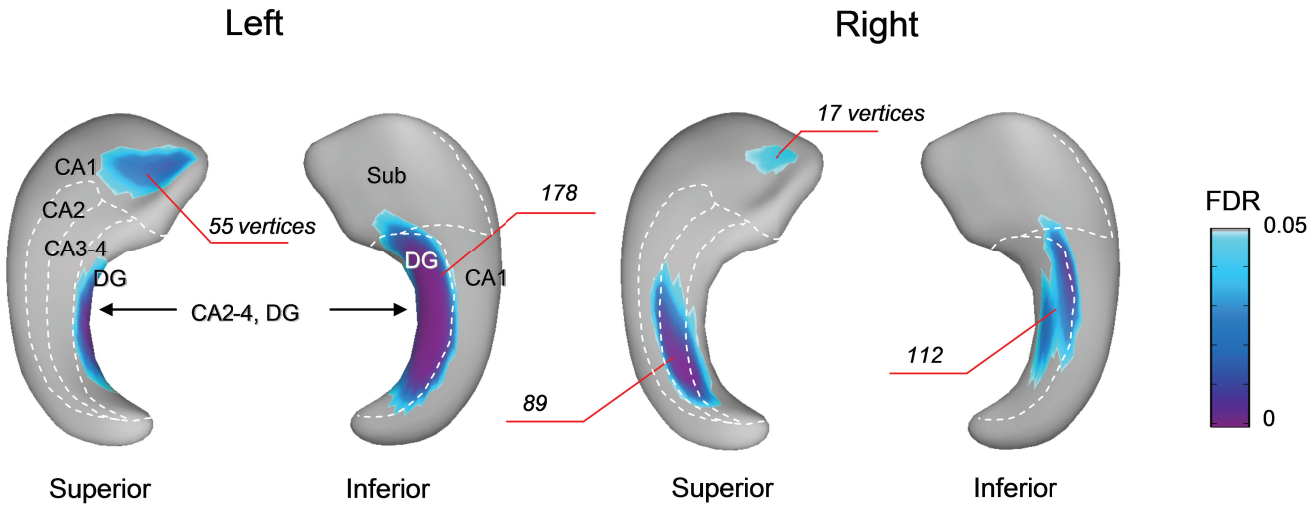


Figure S1—Vertexwise group comparison performed separately in each hemisphere. Pattern of atrophy was very similar between hemispheres and to that found when hemispheres were pooled (Figure 3).

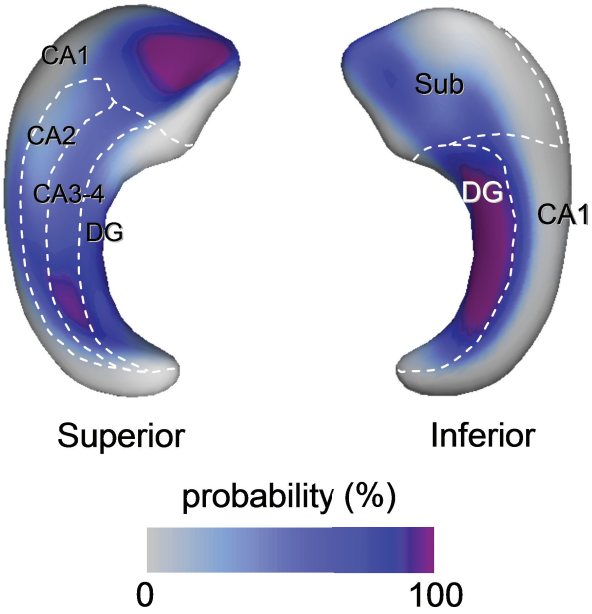


Figure S2—Reproducibility of our surface-based analytic framework. Bootstrap reproducibility analysis revealed that the proposed method highly reproduced the pattern of atrophy in Figure 3 as the same pattern was found in more than 85% of the random subsampling.

Downloaded from https://academic.oup.com/sleep/article-abstract/37/7/1189/2709312
by guest
on 01 February 2018

EXTENDED PDF FORMAT
SPONSORED BY
IMMUNOLOGY2013™ Travel Grants
from R&D Systems
$1000 travel grants will be awarded to 10 immunology researchers
Apply Now!
www.rndsystems.com



Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene

Avshalom Caspi *et al.*

Science **301**, 386 (2003);
DOI: 10.1126/science.1083968

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of January 13, 2013):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/301/5631/386.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2003/07/16/301.5631.386.DC1.html

A list of selected additional articles on the Science Web sites **related to this article** can be found at:
http://www.sciencemag.org/content/301/5631/386.full.html#related

This article **cites 24 articles**, 8 of which can be accessed free:
http://www.sciencemag.org/content/301/5631/386.full.html#ref-list-1

This article has been **cited by** 2022 article(s) on the ISI Web of Science

This article has been **cited by** 100 articles hosted by HighWire Press; see:
http://www.sciencemag.org/content/301/5631/386.full.html#related-urls

This article appears in the following **subject collections:**
Psychology
http://www.sciencemag.org/cgi/collection/psychology

Downloaded from www.sciencemag.org on January 13, 2013

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2003 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

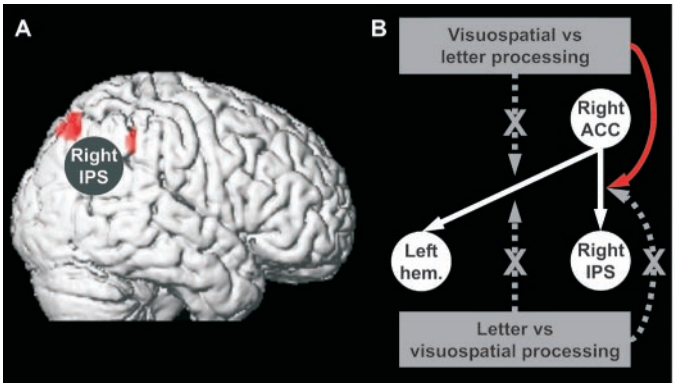


Fig. 4. PPI of the right ACC. (**A**) All areas are shown that receive a significant context-dependent contribution from right ACC during visuospatial decisions, projected on the same rendered brain as in Fig. 1. Right ACC significantly increased its influence on posterior (28/–72/48, $t_{max} = 4.49$, $P < 0.015$, corrected) and anterior parts of the right IPS (42/–42/44, $t_{max} = 4.63$, $P < 0.034$, corrected) during visuospatial decisions. Note the specificity of this result: Even when the threshold was reduced to $P < 0.05$, uncorrected, no other significant clusters were found throughout the brain. (**B**) This schema summarizes the negative findings for right ACC: As indicated by the gray dashed lines, right ACC shows no context-dependent contributions to any left-hemispheric area during visuospatial decisions and none to any left- or right-hemispheric area at all during letter decisions.

tivity (*25*) that investigated two tasks with right hemisphere dominance demonstrated top-down effects that were specific for the right hemisphere, i.e., from the right middle frontal gyrus (area 46) on right extrastriate areas. It is thus likely that our findings generalize to other lateralized tasks. Although we cannot exclude lateralization contingent on stimulus type in some situations, our results are consistent with previous findings from split-brain patient studies (*7*) and positron emission tomography (*11*) showing hemispheric specialization based on task demands. Research on hemispheric specialization should move beyond analyses of asymmetric regional activations and focus more strongly on functional interactions within and between hemispheres.

References and Notes

1. J. L. Bradshaw, N. C. Nettleton, *Behav. Brain Sci.* **4**, 51 (1981).
2. J. B. Hellige, *Annu. Rev. Psychol.* **41**, 55 (1990).
3. J. C. Marshall, *Behav. Brain Sci.* **4**, 72 (1981).
4. K. Hugdahl, R. J. Davidson, *The Asymmetrical Brain* (MIT Press, Cambridge, MA, 2003).
5. J. Sergent, *Psychol. Bull.* **93**, 481 (1983).
6. C. Chiarello, J. Senehi, M. Soulier, *Neuropsychologia* **4**, 521 (1986).
7. P. M. Corballis, M. G. Funnell, M. S. Gazzaniga, *NeuroReport* **10**, 2183 (1999).
8. C. J. Price, R. J. S. Wise, R. S. J. Frackowiak, *Cereb. Cortex* **6**, 62 (1996).
9. G. R. Fink *et al.*, *Nature* **382**, 626 (1996).
10. L. E. Nystrom *et al.*, *Neuroimage* **11**, 424 (2000).
11. S. M. Courtney, L. G. Ungerleider, K. Keil, J. V. Haxby, *Cereb. Cortex* **6**, 39 (1996).
12. J. Levy, C. Trevarthen, *J. Exp. Psychol. Hum. Percept. Perform.* **2**, 299 (1976).
13. M. S. Gazzaniga, *Brain* **123**, 1293 (2000).
14. M. I. Posner, S. E. Petersen, *Annu. Rev. Neurosci.* **13**, 25 (1990).
15. T. Shallice, *From Neuropsychology to Mental Structure* (Cambridge Univ. Press, Cambridge, 1988).
16. R. Desimone, J. Duncan, *Annu. Rev. Neurosci.* **18**, 193 (1995).
17. Materials and methods are available as supporting material on *Science* Online.
18. Using a 2 × 2 × 2 factorial design, with letter vs. visuospatial decisions, left vs. right visual fields of presentation, and left vs. right response hand as the three experimental factors, the eight conditions occurred equally often and were varied systematically as blocked conditions in a pseudorandom fashion.
19. A. W. MacDonald, J. D. Cohen, V. A. Stenger, C. S. Carter, *Science* **288**, 1835 (2000).
20. B. J. Casey *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8728 (2000).
21. C. S. Carter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1944 (2000).
22. G. R. Fink *et al.*, *Brain* **122**, 497 (1999).
23. C. Büchel, K. J. Friston, *Cereb. Cortex* **7**, 768 (1997).
24. A. K. Engel, P. Fries, W. Singer, *Nature Rev. Neurosci.* **2**, 704 (2001).
25. A. R. McIntosh *et al.*, *J. Neurosci.* **14**, 655 (1994).
26. G. Bush, P. Luu, M. I. Posner, *Trends Cogn. Sci.* **4**, 215 (2000).
27. K. J. Friston *et al.*, *NeuroImage* **6**, 218 (1997).
28. C. Cavada, P. S. Goldman-Rakic, *J. Comp. Neurol.* **287**, 393 (1989).
29. D. N. Pandya, G. W. Van Hoesen, M. M. Mesulam, *Exp. Brain Res.* **42**, 319 (1981).

30. This work was supported by the Deutsche Forschungsgemeinschaft (K.Z., G.R.F), the U.K. Medical Research Council (J.C.M), and the Wellcome Trust (K.E.S., K.J.F). Additional support for this Human Brain Project/Neuroinformatics research was provided jointly by the National Institute of Mental Health, National Institute of Neurological Disorders and Stroke, National Institute on Drug Abuse, and the National Cancer Institute. We thank our volunteers, K. Amunts for helpful anatomical discussions, and the radiographers at the Research Center Jülich for technical assistance.

Supporting Online Material
www.sciencemag.org/cgi/content/full/301/5631/384/DC1
Materials and Methods
References

23 April 2003; accepted 6 June 2003

Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene

Avshalom Caspi,[1,2] Karen Sugden,[1] Terrie E. Moffitt,[1,2]* Alan Taylor,[1] Ian W. Craig,[1] HonaLee Harrington,[2] Joseph McClay,[1] Jonathan Mill,[1] Judy Martin,[3] Antony Braithwaite,[4] Richie Poulton[3]

In a prospective-longitudinal study of a representative birth cohort, we tested why stressful experiences lead to depression in some people but not in others. A functional polymorphism in the promoter region of the serotonin transporter (5-HTT) gene was found to moderate the influence of stressful life events on depression. Individuals with one or two copies of the short allele of the 5-HTT promoter polymorphism exhibited more depressive symptoms, diagnosable depression, and suicidality in relation to stressful life events than individuals homozygous for the long allele. This epidemiological study thus provides evidence of a gene-by-environment interaction, in which an individual's response to environmental insults is moderated by his or her genetic makeup.

Depression is among the top five leading causes of disability and disease burden throughout the world (*1*). Across the life span, stressful life events that involve threat, loss, humiliation, or defeat influence the onset and course of depression (*2–5*). However, not all people who encounter a stressful life experience succumb to its depressogenic effect. Diathesis-stress theories of depression predict that individuals' sensitivity to stressful events depends on their genetic makeup (*6*, *7*). Behavioral genetics research supports this prediction, documenting that the risk of depression after a stressful event is elevated among people who are at high genetic risk and diminished among those at low genetic risk (*8*). However, whether specific genes exacerbate or buffer the effect of stressful life events on depression is unknown. In this study, a functional polymorphism in the pro-

[1]Medical Research Council Social, Genetic, and Developmental Psychiatry Research Centre, Institute of Psychiatry, King's College London, PO80 De Crespigny Park, London, SE5 8AF, UK. [2]Department of Psychology, University of Wisconsin, Madison, WI 53706, USA. [3]Dunedin School of Medicine, [4]Department of Pathology, University of Otago, Dunedin, New Zealand.

*To whom correspondence should be addressed. E-mail: t.moffitt@iop.kcl.ac.uk

Downloaded from www.sciencemag.org on January 13, 2013

moter region of the serotonin transporter gene (*SLC6A4*) was used to characterize genetic vulnerability to depression and to test whether 5-HTT gene variation moderates the influence of life stress on depression.

The serotonin system provides a logical source of candidate genes for depression, because this system is the target of selective serotonin reuptake–inhibitor drugs that are effective in treating depression (*9*). The serotonin transporter has received particular attention because it is involved in the reuptake of serotonin at brain synapses (*10*). The promoter activity of the 5-HTT gene, located on 17q11.2, is modified by sequence elements within the proximal 5′ regulatory region, designated the 5-HTT gene-linked polymorphic region (5-HTTLPR). The short ("s") allele in the 5-HTTLPR is associated with lower transcriptional efficiency of the promoter compared with the long ("l") allele (*11*).

Evidence for an association between the short promoter variant and depression is inconclusive (*12*). Although the 5-HTT gene may not be directly associated with depression, it could moderate the serotonergic response to stress. Three lines of experimental research suggest this hypothesis of a gene-by-environment (G × E) interaction. First, in mice with disrupted 5-HTT, homozygous and heterozygous (5-HTT –/– and +/–) strains exhibited more fearful behavior and greater increases in the stress hormone (plasma) adrenocorticotropin in response to stress compared to homozygous (5-HTT +/+) controls, but in the absence of stress no differences related to genotype were observed (*13*). Second, in rhesus macaques, whose length variation of the 5-HTTLPR is analogous to that of humans, the short allele is associated with decreased serotonergic function [lower cerebrospinal fluid (CSF) 5-hydroxyindoleacetic acid concentrations] among monkeys reared in stressful conditions but not among normally reared monkeys (*14*). Third, human neuroimaging research suggests that the stress response is mediated by variations in the 5-HTTLPR. Humans with one or two copies of the s allele exhibit greater amygdala neuronal activity to fearful stimuli compared to individuals homozygous for the l allele (*15*). Taken together, these findings suggest the hypothesis that variations in the 5-HTT gene moderate psychopathological reactions to stressful experiences.

We tested this G × E hypothesis among members of the Dunedin Multidisciplinary Health and Development Study (*16*). This representative birth cohort of 1037 children (52% male) has been assessed at ages 3, 5, 7, 9, 11, 13, 15, 18, and 21 and was virtually intact (96%) at the age of 26 years. A total of 847 Caucasian non-Maori study members, without stratification confounds, were divided into three groups on the basis of their 5-HTTLPR genotype (*11*): those with two copies of the s allele (s/s homozygotes; n = 147; 17%), those with one copy of the s allele (s/l heterozygotes; n = 435; 51%), and those with two copies of the l allele (l/l homozygotes; n = 265; 31%). There was no difference in genotype frequencies between the sexes [$\chi^2(2) = 0.02$, $P = 0.99$]. Stressful life events occurring after the 21st birthday and before the 26th birthday were assessed with the aid of a life-history calendar (*17*), a highly reliable method for ascertaining life-event histories (*18*). The 14 events included employment, financial, housing, health, and relationship stressors. Thirty percent of the study members experienced no stressful life events; 25% experienced one event; 20%, two events; 11%, three events; and 15%, four or more events. There were no significant differences between the three genotype groups in the number of life events they experienced, $F(2,846) = 0.56$, $P = 0.59$, suggesting that 5-HTTLPR genotype did not influence exposure to stressful life events.

Study members were assessed for past-year depression at age 26 with the use of the Diagnostic Interview Schedule (*19*), which yields a quantitative measure of depressive symptoms and a categorical diagnosis of a major depressive episode according to *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) criteria (*20*). 17% of study members (58% female versus 42% male; odds ratio = 1.6; 95% confidence interval from 1.1 to 2.2) met criteria for a past-year major depressive episode, which is comparable to age and sex prevalence rates observed in U.S. epidemiological studies (*21*). In addition, 3% of the study members reported past-year suicide attempts or recurrent thoughts about suicide in the context of a depressive episode. We also collected informant reports about symptoms of depression for 96% of study members at age 26 by mailing a brief questionnaire to persons nominated by each study member as "someone who knows you well."

We used a moderated regression framework (*22*), with sex as a covariate, to test the association between depression and (i) 5-HTTLPR genotype, (ii) stressful life events, and (iii) their interaction (table S1). The interaction between 5-HTTLPR and life events showed that the effect of life events on self-reports of depression symptoms at age 26 was significantly stronger ($P = 0.02$) among individuals carrying an s allele than among l/l homozygotes (Fig. 1A). We further tested whether life events could predict within-individual increases in depression symptoms over time among individuals with an s allele by statistically controlling for the baseline number of depressive symptoms they had before the life events occurred (table S1). The significant interaction ($P = 0.05$) showed that individuals carrying an s allele whose life events occurred after their 21st birthday experienced increases in depressive symptoms from the age of 21 to 26 years ($b = 1.55$, SE = 0.66, $t = 2.35$, $P = 0.02$ among s/s homozygotes and $b = 1.25$, SE = 0.34, $t = 3.66$, $P < 0.001$ among s/l heterozygotes) whereas l/l homozygotes did not ($b = 0.17$, SE = 0.41, $t = 0.41$, $P = 0.68$).

The G × E interaction also showed that stressful life events predicted a diagnosis of major depression among carriers of an s allele but not among l/l homozygotes ($P = 0.056$, Fig. 1B). We further tested whether life events could predict the onset of new diagnosed depression among carriers of an s allele (table S1). We excluded from analysis study members who were diagnosed with depression before age 21. The significant interaction ($P = 0.02$) showed that life events occurring after their 21st birthdays predicted depression at age 26 among carriers of an s allele who did not have a prior history of depression ($b = 0.79$, SE = 0.25, $z = 3.16$, $P = 0.002$ among s/s homozygotes and $b = 0.41$, SE = 0.12, $z = 3.29$, $P = 0.001$ among s/l heterozygotes) but did not predict onset of new depression among l/l homozygotes ($b = 0.08$, SE = 0.20, $z = 0.42$, $P = 0.67$). Further analyses showed that stressful life events predicted suicide ideation or attempt among individuals carrying an s allele but not among l/l homozygotes ($P = 0.05$, Fig. 1C). The hypothesized G × E interaction was also significant when we predicted informant reports of age-26 depression ($P < 0.01$), an analysis that ruled out the possibility of self-report bias (Fig. 1D). The interaction showed that the effect of life events on informant reports of depression was stronger among individuals carrying an s allele than among l/l homozygotes. These analyses attest that the 5-HTT gene interacts with life events to predict depression symptoms, an increase in symptoms, depression diagnoses, new-onset diagnoses, suicidality, and an informant's report of depressed behavior.

This evidence that 5-HTTLPR variation moderates the effect of life events on depression does not constitute unambiguous evidence of a G × E interaction, because exposure to life events may be influenced by genetic factors; if individuals have a heritable tendency to enter situations where they encounter stressful life events, these events may simply be a genetically saturated marker (*23*, *24*). Thus, what we have identified as a gene × environment interaction predicting depression could actually reflect a gene × "gene" interaction between the 5-HTTLPR and other genes we did not measure. We reasoned that, if our measure of life events represents merely genetic risk, then life events would interact with 5-HTTLPR even if they occurred after the depression episode. However, if our measure of life events represents environmental stress, then the timing of life events relative to depression must follow cause-effect order and life events that

Downloaded from www.sciencemag.org on January 13, 2013

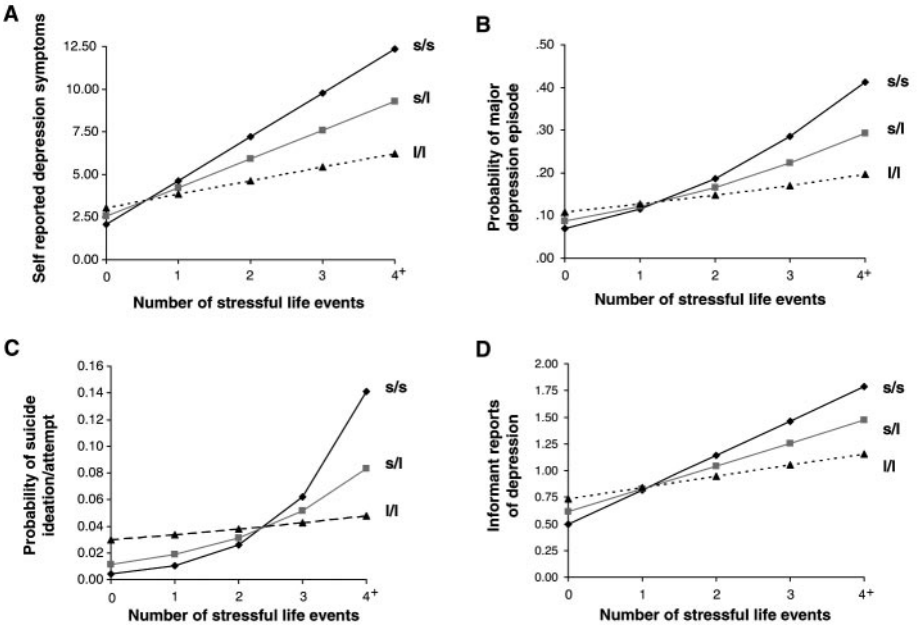


Fig. 1. Results of multiple regression analyses estimating the association between number of stressful life events (between ages 21 and 26 years) and depression outcomes at age 26 as a function of 5-HTT genotype. Among the 146 s/s homozygotes, 43 (29%), 37 (25%), 28 (19%), 15 (10%), and 23 (16%) study members experienced zero, one, two, three, and four or more stressful events, respectively. Among the 435 s/l heterozygotes, 141 (32%), 101 (23%), 76 (17%), 49 (11%), and 68 (16%) experienced zero, one, two, three, and four or more stressful events. Among the 264 l/l homozygotes, 79 (29%), 73 (28%), 57 (21%), 26 (10%), and 29 (11%) experienced zero, one, two, three, and four or more stressful events. (**A**) Self-reports of depression symptoms. The main effect of 5-HTTLPR (i.e., an effect not conditional on other variables) was marginally significant ($b = -0.96$, SE = 0.52, $t = 1.86$, $P = 0.06$), the main effect of stressful life events was significant ($b = 1.75$, SE = 0.23, $t = 7.45$, $P < 0.001$), and the interaction between 5-HTTLPR and life events was in the predicted direction ($b = -0.89$, SE = 0.37, $t = 2.39$, $P = 0.02$). The interaction showed that the effect of life events on self-reports of depression symptoms was stronger among individuals carrying an s allele ($b = 2.52$, SE = 0.66, $t = 3.82$, $P < 0.001$ among s/s homozygotes, and $b = 1.71$, SE = 0.34, $t = 5.02$, $P < 0.001$ among s/l heterozygotes) than among l/l homozygotes ($b = 0.77$, SE = 0.43, $t = 1.79$, $P = 0.08$). (**B**) Probability of major depressive episode. The main effect of 5-HTTLPR was not significant ($b = -0.15$, SE = 0.14, $z = 1.07$, $P = 0.29$), the main effect of life events was significant ($b = 0.37$, SE = 0.06, $z = 5.99$, $P < 0.001$), and the G × E was in the predicted direction ($b = -0.19$, SE = 0.10, $z = 1.91$, $P = 0.056$). Life events predicted a diagnosis of major depression among s carriers ($b = 0.52$, SE = 0.16, $z = 3.28$, $P = 0.001$ among s/s homozygotes, and $b = 0.39$, SE = 0.09, $z = 4.24$, $P < 0.001$ among s/l heterozygotes) but not among l/l homozygotes ($b = 0.16$, SE = 0.13, $z = 1.18$, $P = 0.24$). (**C**) Probability of suicide ideation or attempt. The main effect of 5-HTTLPR was not significant ($b = -0.01$, SE = 0.28, $z = 0.01$, $P = 0.99$), the main effect of life events was significant ($b = 0.51$, SE = 0.13, $z = 3.96$, $P < 0.001$), and the G × E interaction was in the predicted direction ($b = -0.39$, SE = 0.20, $t = 1.95$, $P = 0.051$). Life events predicted suicide ideation or attempt among s carriers ($b = 0.48$, SE = 0.29, $z = 1.67$, $P = 0.09$ among s/s homozygotes, and $b = 0.91$, SE = 0.25, $z = 3.58$, $P < 0.001$ among s/l heterozygotes) but not among l/l homozygotes ($b = 0.13$, SE = 0.26, $z = 0.49$, $P = 0.62$). (**D**) Informant reports of depression. The main effect of 5-HTTLPR was not significant (b = −0.06, SE = 0.06, t = 0.98, $P = 0.33$), the main effect of life events was significant ($b = 0.23$, SE = 0.03, $t = 8.47$, $P < 0.001$), and the G × E was in the predicted direction ($b = -0.11$, SE = 0.04, $t = 2.54$, $P < 0.01$). The effect of life events on depression was stronger among s carriers ($b = 0.39$, SE = 0.07, $t = 5.23$, $P < 0.001$ among s/s homozygotes, and $b = 0.17$, SE = 0.04, $t = 4.51$, $P < 0.001$ among s/l heterozygotes) than among l/l homozygotes ($b = 0.14$, SE = 0.05, $t = 2.69$, $P < 0.01$).

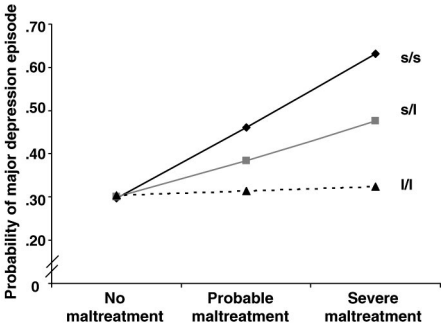


Fig. 2. Results of regression analysis estimating the association between childhood maltreatment (between the ages of 3 and 11 years) and adult depression (ages 18 to 26), as a function of 5-HTT genotype. Among the 147 s/s homozygotes, 92 (63%), 39 (27%), and 16 (11%) study members were in the no maltreatment, probable maltreatment, and severe maltreatment groups, respectively. Among the 435 s/l heterozygotes, 286 (66%), 116 (27%), and 33 (8%) were in the no, probable, and severe maltreatment groups. Among the 265 l/l homozygotes, 172 (65%), 69 (26%), and 24 (9%) were in the no, probable, and severe maltreatment groups. The main effect of 5-HTTLPR was not significant ($b = -0.14$, SE = 0.11, $z = 1.33$, $P = 0.19$), the main effect of childhood maltreatment was significant ($b = 0.30$, SE = 0.10, $z = 3.04$, $P = 0.002$), and the G × E interaction was in the predicted direction ($b = -0.33$, SE = 0.16, $z = 2.01$, $P = 0.05$). The interaction showed that childhood stress predicted adult depression only among individuals carrying an s allele ($b = 0.60$, SE = 0.26, $z = 2.31$, $P = 0.02$ among s/s homozygotes, and $b = 0.45$, SE = 0.16, $z = 2.83$, $P = 0.01$ among s/l heterozyotes) and not among l/l homozygotes ($b = -0.01$, SE = 0.21, $z = 0.01$, $P = 0.99$).

occur after depression should not interact with 5-HTTLPR to postdict depression. We tested this hypothesis by substituting the age-26 measure of depression with depression assessed in this longitudinal study when study members were 21 and 18 years old, before the occurrence of the measured life events between the ages of 21 and 26 years. Whereas the 5-HTTLPR × life events interaction predicted depression at the age of 26 years, this same interaction did not postdict depression reported at age 21 nor at the age of 18 years (table S2), indicating our finding is a true G × E interaction.

If 5-HTT genotype moderates the depressogenic influence of stressful life events, it should moderate the effect of life events that occurred not just in adulthood but also of stressful experiences that occurred in earlier developmental periods. Based on this hypothesis, we tested whether adult depression was predicted by the interaction between 5-HTTLPR and childhood maltreatment that occurred during the first decade of life (*16*, *25*). Consistent with the G × E hypothesis, the longitudinal prediction from childhood maltreatment to adult depression was significantly moderated by 5-HTTLPR (table S3). The interaction showed ($P = 0.05$) that childhood maltreatment predicted adult depression only among individuals carrying an s allele but not among l/l homozygotes (Fig. 2).

We previously showed that variations in the gene encoding the neurotransmitter-metabolizing enzyme monoamine oxidase A (MAOA) moderate children's sensitivity to maltreatment (*25*). MAOA has high affinity for 5-HTT, raising the possibility that the protective effect of the l/l allele on psychiatric morbidity is further augmented by the presence of a genotype conferring high MAOA activity (*13*, *26*). However, we found that the moderation of life stress on depression was specific to a polymorphism in the 5-HTT gene, because this effect was observed regardless of the individual's MAOA gene status (tables S4 and S5).

Until this study's findings are replicated, speculation about clinical implications is pre-

Downloaded from www.sciencemag.org on January 13, 2013

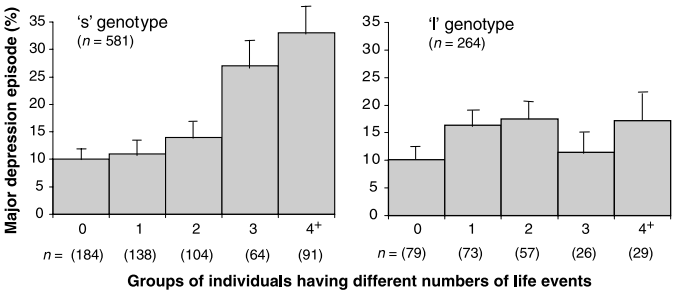


Fig. 3. The percentage of individuals meeting diagnostic criteria for depression at age 26, as a function of 5-HTT genotype and number of stressful life events between the ages of 21 and 26. The figure shows individuals with either one or two copies of the short allele (**left**) and individuals homozygous for the long allele (**right**). In a hierarchical logistic regression model, the main effect of genotype (coded as s group = 0 and l group = 1) was not significant, $b = -0.15$, SE = 0.21, $z = 0.72$, $P = 0.47$; the main effect of number of life events was significant, $b = 0.34$, SE = 0.06, $z = 5.70$, $P < 0.001$; and the interaction between genotype and number of life events was significant, $b = -0.30$, SE = 0.15, $z = 1.97$, $P = 0.05$.

mature. Nonetheless, although carriers of an s 5-HTTLPR allele who experienced four or more life events constituted only 10% of the birth cohort, they accounted for almost one-quarter (23%) of the 133 cases of diagnosed depression. Moreover, among cohort members suffering four or more stressful life events, 33% of individuals with an s allele became depressed, whereas only 17% of the l/l homozygotes developed depression (Fig. 3). Thus, the G × E's attributable risk and predictive sensitivity indicate that more knowledge about the functional properties of the 5-HTT gene may lead to better pharmacological treatments for those already depressed. Although the short 5-HTTLPR variant is too prevalent for discriminatory screening (over half of the Caucasian population has an s allele), a microarray of genes might eventually identify those needing prophylaxis against life's stressful events (*27*).

Evidence of a direct relation between the 5-HTTLPR and depression has been inconsistent (*12*), perhaps because prior studies have not considered participants' stress histories. In this study, no direct association between the 5-HTT gene and depression was observed. Previous experimental paradigms, including 5-HTT knockout mice (*13*), stress-reared rhesus macaques (*14*), and human functional neuroimaging (*15*), have shown that the 5-HTT gene can interact with environmental conditions, although these experiments did not address depression. Our study demonstrates that this G × E interaction extends to the natural development of depression in a representative sample of humans. However, we could not test hypotheses about brain endophenotypes (*28*) intermediate between the 5-HTT gene and depression because of the difficulty of taking CSF or functional magnetic resonance imaging measurements in an epidemiological cohort.

Much genetic research has been guided by the assumption that genes cause diseases, but the expectation that direct paths will be found from gene to disease has not proven fruitful for complex psychiatric disorders (*29*). Our findings of G × E interaction for the 5-HTT gene and another candidate gene, MAOA (*25*), point to a different, evolutionary model. This model assumes that genetic variants maintained at high prevalence in the population probably act to promote organisms' resistance to environmental pathogens (*30*). We extend the concept of environmental pathogens to include traumatic, stressful life experiences and propose that the effects of genes may be uncovered when such pathogens are measured (in naturalistic studies) or manipulated (in experimental studies). To date, few linkage studies detect genes, many candidate gene studies fail consistent replication, and genes that replicate account for little variation in the phenotype (*29*). If replicated, our G × E findings will have implications for improving research in psychiatric genetics. Incomplete gene penetrance, a major source of error in linkage pedigrees, can be explained if a gene's effects are expressed only among family members exposed to environmental risk. If risk exposure differs between samples, candidate genes may fail replication. If risk exposure differs among participants within a sample, genes may account for little variation in the phenotype. We speculate that some multifactorial disorders, instead of resulting from variations in many genes of small effect, may result from variations in fewer genes whose effects are conditional on exposure to environmental risks.

References and Notes

1. C. J. Tang, A. D. Lopez, *Lancet* **349**, 1498 (1997).
2. G. W. Brown, *Soc. Psychiatry Psychiatr. Epidemiol.* **33**, 363 (1998).
3. K. S. Kendler, L. M. Karkowski, C. A. Prescott, *Am. J. Psychiatry* **156**, 837 (1999).
4. R. C. Kessler, *Annu. Rev. Psychol.* **48**, 191 (1997).
5. D. S. Pine, P. Cohen, J. G. Johnson, J. S. Brook, *J. Affect. Disorders* **68**, 49 (2002).
6. E. J. Costello *et al.*, *Biol. Psychiatry* **52**, 529 (2002).
7. S. M. Monroe, A. D. Simons, *Psychol. Bull.* **110**, 406 (1991).
8. K. S. Kendler *et al.*, *Am. J. Psychiatry* **152**, 833 (1995).
9. C. A. Tamminga *et al.*, *Biol. Psychiatry* **52**, 589 (2002).
10. K. P. Lesch, M. D. Greenberg, J. D. Higley, A. Bennett, D. L. Murphy, in *Molecular Genetics and the Human Personality*, J. Benjamin, R. P. Ebstein, R. H. Belmaker, Eds. [American Psychiatric Association (APA), Washington, DC, 2002], pp. 109–136.
11. K. P. Lesch *et al.*, *Science* **274**, 1527 (1996).
12. K. P. Lesch, in *Behavioral Genetics in the Postgenomics Era*, R. Plomin, J. C. DeFries, I. W. Craig, P. McGuffin, Eds. (APA, Washington, DC, 2003), pp. 389–424.
13. D. L. Murphy *et al.*, *Brain Res. Bull.* **56**, 487 (2001).
14. A. J. Bennett *et al.*, *Mol. Psychiatry* **7**, 188 (2002).
15. A. R. Hariri *et al.*, *Science* **297**, 400 (2002).
16. Materials and methods are available as supporting material on *Science* Online.
17. A. Caspi *et al.*, *Int. J. Methods Psychiatr. Res.* **6**, 101 (1996).
18. R. F. Belli, W. L. Shay, F. P. Stafford, *Public Opin. Q.* **65**, 45 (2001).
19. L. N. Robins, L. Cottler, K. Bucholtz, W. Compton, *Diagnostic Interview Schedule for DSM-IV* (Washington University, St. Louis, MO, 1995).
20. APA, *Diagnostic and Statistical Manual of Mental Disorders* (APA, Washington, DC, ed. 4, 1994).
21. R. C. Kessler, K. A. McGonagle, M. Swartz, D. G. Blazer, C. B. Nelson, *J. Affect. Disorders* **29**, 85 (1993).
22. L. S. Aiken, S. G. West, *Multiple Regression: Testing and Interpreting Interactions* (Sage, Thousand Oaks, CA, 1991).
23. K. S. Kendler, L. Karkowski-Shuman, *Psychol. Med.* **27**, 539 (1997).
24. R. Plomin, C. S. Bergeman, *Behav. Brain Sci.* **14**, 373 (1991).
25. A. Caspi *et al.*, *Science* **297**, 851 (2002).
26. N. Salichon *et al.*, *J. Neurosci.* **21**, 884 (2001).
27. W. E. Evans, M. V. Relling, *Science* **286**, 487 (1999).
28. I. I. Gottesman, T. D. Gould, *Am. J. Psychiatry* **160**, 636 (2003).
29. D. Hamer, *Science* **298**, 71 (2002).
30. A. V. S. Hill, *Br. Med. Bull.* **55**, 401 (1999).

31. We thank P. Silva, founder of the Dunedin Multidisciplinary Health and Development Study, Air New Zealand, and the study members, their families, and their friends. Supported by the Health Research Council of New Zealand and the University of Wisconsin Graduate School and by grants from the U.K. Medical Research Council, the William T. Grant Foundation, and the U.S. National Institute of Mental Health (MH49414 and MH45070). T.E.M. is a Royal Society–Wolfson Research Merit Award holder. The study protocol was approved by the institutional review boards of the participating universities.

Supporting Online Material

www.sciencemag.org/cgi/content/full/301/5631/386/DC1
Materials and Methods
Tables S1 to S5

27 February 2003; accepted 16 June 2003

Downloaded from www.sciencemag.org on January 13, 2013

American College of Neuropsychopharmacology

www.nature.com/npp

ARTICLE OPEN

Check for updates

Infralimbic BDNF signaling is necessary for the beneficial effects of extinction on set shifting in stressed rats

Denisse Paredes[1], Anna R. Knippenberg[1] and David A. Morilak[1,2]✉

© The Author(s) 2021

Current pharmacotherapies for posttraumatic stress disorder (PTSD) and major depressive disorder (MDD) are ineffective for many patients, and often do not restore cognitive dysfunction associated with these disorders. Behavioral therapies, such as exposure therapy, can be effective for treatment-resistant patients. The mechanisms underlying exposure therapy are not well-understood. Fear extinction as an intervention after chronic stress can model the beneficial effects of exposure therapy in rats. Extinction requires neuronal activity and protein synthesis in the infralimbic (IL) cortex for its beneficial effects. We hypothesized that extinction requires Brain-Derived Neurotrophic Factor (BDNF) activity in the IL cortex to reverse stress-induced cognitive flexibility impairments. Extinction learning reversed set-shifting deficits induced by Chronic Unpredictable Stress (CUS), tested 24 h after extinction. Blocking BDNF signaling in the IL cortex during extinction by local administration of a neutralizing antibody prevented the beneficial effects of extinction on set shifting after stress. Extinction induced activation of the BDNF TrkB receptor, and signaling pathways associated with BDNF (Akt and Erk). Administration of exogenous BDNF into IL cortex in the absence of extinction was sufficient to reverse the effects of stress on set shifting. The effects of extinction were prevented by blocking either Erk or Akt signaling in the IL cortex, whereas the effects of exogenous BDNF were dependent on Erk, but not Akt, signaling. Our observations suggest that BDNF-Erk signaling induced by extinction underlies plastic changes that can reverse or counteract the effects of chronic stress in the IL cortex.

Neuropsychopharmacology; https://doi.org/10.1038/s41386-021-01171-7

INTRODUCTION

Psychiatric disorders such as posttraumatic stress disorder (PTSD) and major depressive disorder (MDD), are often comorbid, and affect millions of people every year [1, 2]. These disorders share pathological features, such as maladaptive shifts in coping behavior, and deficits in executive function. Selective serotonin reuptake inhibitors (SSRIs) are the most commonly prescribed pharmacotherapy for such disorders, but only approximately 60% of patients respond to treatment. Behavioral therapies, such as exposure therapy, can be effective [3]. However, only a small percentage of patients achieve full remission after exposure therapy [4]. Deciphering the neurobiological mechanisms underlying exposure therapy could lead to the development of strategies to improve therapeutic outcomes.

Cognitive flexibility is the ability to modify previously established associations, thoughts, or responses based on changes in the environment [30]. Cognitive flexibility deficits represent a symptom dimension shared across several psychiatric disorders, are associated with increased symptom severity in PTSD [5], and are a risk factor for developing MDD [6]. Set shifting is a form of cognitive flexibility that relies on the function of the medial prefrontal cortex (mPFC) [7], a brain region susceptible to the detrimental effects of stress [8]. Set shifting is a change in response strategy (i.e., a "cognitive set") that requires one to redirect attention from one perceptual dimension to another dimension previously irrelevant to the strategy [7]. Behavioral therapies, such as exposure therapy, address underlying cognitive dysfunction instead of targeting individual symptoms that may result from such dysfunction [9]. Exposure therapy has been shown to improve set shifting [10]. Behavioral therapies engage areas of the brain, such as the hippocampus, prefrontal cortex, and amygdala, that are affected by chronic stress [11].

Exposure therapy involves modifying one's cognitive appraisal of fear by utilizing fear extinction (FE) learning, a form of safety learning that consists of forming a new association between a conditioned stimulus and conditioned response [12, 13]. FE in rodents mimics some of the beneficial effects of exposure therapy [14, 15]. During FE, rodents learn that a previously conditioned stimulus (e.g., a tone) no longer signals an aversive stimulus (e.g., footshock). The resulting safety memory is formed through activity in the infralimbic (IL) portion of mPFC [16]. We have shown that extinction is effective in reversing stress-induced deficits in set shifting [14]. Extinction restores responsivity of the mPFC that is compromised by chronic stress and promotes adaptive active coping behavior, similar to the effects of exposure therapy in humans [17, 18]. These effects of extinction require activity-dependent protein translation in the IL [17]. However, the molecular mechanisms that modulate IL plasticity during extinction are not known.

Brain-Derived Neurotrophic Factor (BDNF) is necessary for extinction learning and consolidation in rodents and humans alike [19, 20]. Additionally, exogenous administration of BDNF into

[1]Department of Pharmacology and Center for Biomedical Neuroscience, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA. [2]South Texas Veterans Health Care System, San Antonio, TX, USA. ✉email: Morilak@uthscsa.edu

Received: 10 March 2021 Revised: 24 August 2021 Accepted: 27 August 2021
Published online: 08 September 2021

the IL in the absence of extinction reduces fear (i.e., freezing) to the conditioned stimulus, mimicking extinction [21]. Therefore, we hypothesized that BDNF signaling in the IL may also be involved in the beneficial effects of extinction on set shifting after stress. BDNF can initiate plasticity through phosphorylation of its receptor, Tropomyosin receptor kinase B (TrkB), at tyrosine residues Y515 and Y816. Phosphorylation of Y515 activates the mitogen-activated protein kinase (MAPK)/extracellular regulated protein kinase (Erk) and phosphatidylinositol 3-kinase (PI3k)/protein kinase B (Akt) signaling pathways, both of which modulate neuronal plasticity and protein translation, and have been associated with extinction memory consolidation [22, 23], whereas TrkB phosphorylation at Y816 activates phospholipase C (PLC) signaling [24]. Dysregulation of both Erk and Akt signaling has been implicated in psychiatric disorders [25]. Erk and Akt are reduced in the mPFC and hippocampus of depressed and suicidal individuals, and studies report increases in Erk after antidepressant treatment [25]. Additionally, glucocorticoid exposure decreases phosphorylation of Erk in the hippocampus and increases immobility on the forced swim test [26]. We hypothesized that BDNF activity in the IL during extinction may be necessary to induce functional changes in stressed animals as a result of learning-induced activation of MAPK/Erk and/or PI3k/Akt signaling. Portions of this work have been presented in abstract form [27].

MATERIALS & METHODS

Animals

A total of 258 male and naturally cycling female rats (Envigo, 225–249 g) were group-housed on a 12/12 hr light cycle (lights on at 07:00 hr), with same-sex cagemates upon arrival. After 1 week of acclimation, they were single-housed and provided with food and water *ad libitum*. Experiments took place between 08:00-17:00 hr, during the lights-on portion of the cycle. All procedures were in accordance with NIH guidelines and approved by the UTHSA Institutional Animal Care and Use Committee.

Fear conditioning and extinction

Two days prior to the start of Chronic Unpredictable Stress (CUS) procedures, rats were habituated to two contexts in sound-attenuating cabinets for 15 min each. Context A consisted of a metal conditioning chamber (30.5 × 25.4 × 30.5 cm; model H10-11RC-TC-Sf) with square walls and a grid floor attached to a shock box (model H13–15). Context B was not associated with a shock and was in a different chamber, with pink and white vinyl floor and white vinyl circular walls.

Day 0 - Fear conditioning

Rats received fear conditioning or tone control treatment in Context A the day before the start of CUS. Fear conditioning consisted of four pairings of a tone (10 kHz, 75 dB, 20 s) coterminus with a footshock (0.8 mA, 0.5 s), with an inter-trial interval of 120 s. Freezing during each tone was quantified videographically by FreezeView software (ActiMetrics #ACT-100, Coulbourn Instruments). Tone-control rats were exposed to the tones but did not receive a shock.

Day 17 - Fear extinction

Three days after the end of CUS, all rats received a single fear extinction session in Context B. All rats were exposed to 16 tone-alone trials with no shock, with an average inter-trial interval of 120 s.

Chronic Unpredictable stress (CUS)

Chronic unpredictable stress was administered as previously described [17, 28]. For males, CUS consisted of 14 days of varied acute stressors: 30-min restraint, 1 h shake, 45 min social defeat, 10 min tail pinch, 24 h wet bedding, and 15 min mild footshock. We have reported that 14 days of CUS is not sufficient to induce set shifting deficits in females, so 21-day CUS was used for females, which induces the same set-shifting deficit as 14-day CUS in males [28]. CUS for females entailed mild footshock, tail pinch, shake stress, restraint, 24 hr wet bedding, and overnight lights on in place of social defeat.

Attentional set shifting test (AST)

The AST was used to measure cognitive flexibility on the extra-dimensional set shifting task as previously described [29]. A week prior to testing, animals were food restricted to 12 g/day for males and 9 g/day for females (~66% of average daily intake). Animals were trained and tested in a white plastic arena containing a start gate at one end and two terracotta pots separated by a Plexiglas wall at the other end. The AST consists of three days (habituation, training, and testing), with an extra day inserted between training and testing to allow for the administration of extinction:

i.Habituation day: Rats were trained to dig in pots filled with sawdust to retrieve a food reward, half a Honey Nut Cheerio (General Mills Cereals, Minneapolis, MN, USA).

ii. Training day: Rats were trained to locate the reward in one of two pots by discriminating cues in two stimulus dimensions: the odor applied to the rim of the pot or the texture of the digging medium that filled the pot.

iii. Treatment day: Rats received bilateral microinjections and extinction or tone control treatment.

iv. Testing day: The day after extinction, rats were tested on a series of discrimination tasks, in which a criterion of six correct consecutive trials was required to proceed to the next task (Table 1). Rats were tested using either medium or odor as the relevant dimension in the early discrimination tasks (i.e., simple discrimination, complex discrimination, reversal learning, intradimensional shift, second reversal), leading to the formation of a cognitive set. In the extra-dimensional (ED) set shifting task, the previously relevant dimension (e.g., odor), was irrelevant while the previously irrelevant dimension (medium) now indicated the location of the reward.

Western blotting

17 days after rats were fear conditioned or received tone control treatment, rats underwent fear extinction and 30 min later were sacrificed via rapid decapitation. In experiments where animals were stressed, rats were fear conditioned on day 0, then received stress or handling from days 1–14 for males or 1–21 days for females. On days 15-16 (days 22-23 for females), rats were left undisturbed to coincide with the experimental timeline of the AST experiments. On day 17 (24 for females), rats received fear extinction and were sacrificed 30 min after the end of extinction. The IL was dissected on ice from a 2 mm coronal slab cut 2–4 mm caudal to the frontal pole. Cortex adjacent and medial to the internal capsule was dissected from this slab, frozen in isopentane on dry ice, and stored at −80 °C until use. In preliminary experiments measuring phosphorylation of Y515 in IL at 0′, 30′, and 60′ post extinction, we found an increase in phosphorylation of Y515 at 30 min post extinction. Thus, we selected 30′ as our tissue collection time. Western blots were performed as described previously [30]. Membranes were incubated in anti-rabbit or mouse secondary antibody (1:5,000 Cell Signaling) detected using ECL Prime (GE Healthcare, Little Chalfont, UK). Membranes were stripped and re-probed with antibodies against total TrkB (1:1000, Neuromics, [31]), ERK (rabbit pAb 1:5000, Santa Cruz Biotechnology, sc-94), or Akt (mouse mAb 1:1,000, CST 29202). Images were captured using the G:BOX-XT4 Chemi system (Syngene; Frederick,MD). All experimental group samples were normalized to the mean of tone control samples run in the same assay.

Table 1. Stages of the Attentional Set Shifting Test (AST). Order of discrimination stages used for AST. All data shown are from the extra dimensional (ED) set shifting task.

	Dimensions		Example combinations	
Discrimination stage	Relevant	Irrelevant	(+)	(−)
Simple	Odor		Clove/sawdust	Nutmeg/sawdust
Compound	Odor	Medium	Clove/raffia	Nutmeg/yarn
			Clove/yarn	Nutmeg/raffia
Reversal 1	Odor	Medium	Nutmeg/raffia	Clove/yarn
			Nutmeg/yarn	Clove/raffia
Intradimensional shift	Odor	Medium	Rosemary/wood balls	Cinnamon/plastic beads
			Rosemary/plastic beads	Cinnamon/wood balls
Reversal 2	Odor	Medium	Cinnamon/plastic beads	Rosemary/wood balls
			Cinnamon/ wood balls	Rosemary/plastic beads
Extradimensional shift (ED)	Medium	Odor	Velvet/citronella	Crepe/thyme
			Velvet/thyme	Crepe/citronella

Extinction with microinjections of BDNF neutralizing antibody and signaling pathway inhibitors

Rats were implanted with microinjection guide cannulae (Plastics One, Roanoke, VA) positioned 1 mm above the IL (AP + 2.9, ML −3.1, DV −3.8) at a 30° angle to minimize disruption of the prelimbic cortex. After the completion of experiments, a subset of rat brains from each group was used to verify cannula placement. Animals recovered for 1 week after surgery. On day 0, they received fear conditioning or tone exposure and on days 1–14 (1–21 for females), they underwent CUS. In this experiment and the next, in which exogenous BDNF was administered (below), males received overnight lights on in place of social defeat to make their CUS protocol identical to that of females except for the duration. There was no difference in the effect of CUS. On days 15–16, they were habituated and trained on the AST. On day 17, they received bilateral microinjection prior to extinction training. Infusion cannulae were inserted, extending 1 mm beyond the guide cannulae. Rats received bilateral injections of sheep IgG or sheep anti-BDNF (EMD Millipore, Billerica, MA), 0.5 µg/0.5 µl per side at a rate of 0.1 µl/min prior to extinction. This dose was shown to block BDNF signaling [24]. The injector remained in place for 2 min for diffusion. Twenty minutes after completing microinjections, rats underwent extinction. 24 h following extinction and microinjections, rats were tested on the set shifting test. For experiments involving kinase inhibitors, rats received bilateral microinjections of the PI3k inhibitor LY294002 (4.3 ng/0.5 µl/side), the Erk inhibitor PD98059 (2 µg/0.5 µl/side), or vehicle. As the different vehicles (0.5% DMSO in saline, 70% DMSO in saline, or saline alone) did not differ and had no effect on behavior, these groups were combined into a single vehicle group to minimize animal usage. Both LY294002 and PD98059 at the selected doses effectively inhibit phosphorylation of Akt and Erk in vivo, respectively [30, 32]. Based on preliminary western blot experiments showing phosphorylation of TrkB at Y515 occurred 30 mins post FE, we administered the inhibitors immediately after the end of the 32-minute extinction session so drug administration would precede induction of phosphorylation of Y515. One day later, rats were tested on the AST.

Microinjections of BDNF peptide and signaling pathway inhibitors in place of extinction

Rats were implanted with microinjection guide cannulae (Plastics One, Roanoke, VA) positioned 1 mm above the IL (AP + 2.9, ML −3.1, DV −3.8) at a 30° angle. Animals recovered for one week after surgery. On days 1–14 for males and days 1–21 for females, rats underwent CUS. On days 15-16 (22–23 for females) they were habituated and trained on the AST. On day 17 (24 for females), they received bilateral microinjections (0.75 µl/side) into IL of BDNF alone (0.5 µg/µl, R&D Systems), or a combination of BDNF after either the PI3k inhibitor LY294002 (4.3 ng/0.5 µl/side), the Erk inhibitor PD98059 (2 µg/0.5 µl/side), or vehicle (0.5% DMSO in 0.9% saline or saline). The inhibitors were injected first, followed by BDNF. The microinjector remained in place for 2 min to allow diffusion. 24 h after microinjections, rats were tested on the set shifting test.

RESULTS

The BDNF receptor TrkB is phosphorylated in IL at Y515 but not Y816 after fear extinction

32 rats were used in 2 groups (tone controls vs. extinction), subsets of which were used to assess phosphorylation of TrkB at Y515 (9–10 males, 5–8 females), and Y816 (7–8 males, 4–5 females) in the IL cortex. In the aggregate analysis, extinction increased phosphorylation of TrkB at Y515 compared to tone controls in the IL 30 min after extinction ($t_{30} = 3.432$, $p = 0.0018$, Fig. 1A). Broken out by sex, a student's *t*-test showed an effect of extinction on Y515 in males ($t_{17} = 3.568$, $p = 0.0024$). In females, the effect of extinction on Y515 phosphorylation was not significant, although these experiments were not specifically powered for separate analyses by sex ($t_{11} = 1.222$, $p = 0.2474$). Extinction did not induce an increase at Y816 ($t_{22} = 0.4897$, $p = 0.6292$; Fig. 1B). There was no effect of extinction on Y816 in either males (7-8/group; $t_{13} = 0.00213$, $p = 0.9983$, Fig. 1B, top inset) or females (4–5/group; $t_7 = 0.6168$, $p = 0.5569$; Fig. 1B, bottom inset). Therefore, 30 min after extinction, BDNF-TrkB signaling is initiated in the IL by phosphorylation of Y515 but not Y816. Fig. 1C shows the typical pattern of freezing behavior during presentation of the tone during fear conditioning and extinction.

The BDNF receptor TrkB is phosphorylated at Y515 in the IL of both stressed and control animals after the extinction

To test whether extinction also increases phosphorylation at Y515 in stressed rats, 35 rats (18 males, 4–9/group; 17 females, 4–5/group) were used in 4 groups (stress or unstressed x extinction or tone controls) for western blotting. There was no difference in extinction between stress and control groups ($F_{1,16} = 0.03366$, $p = 0.8567$, Fig. 2A). ANOVA revealed a main effect of extinction, indicating an increase in phosphorylation of Y515 in IL ($F_{1,31} = 5.073$, $p = 0.0315$). There was no interaction ($F_{1,31} = 0.9371$, $p = 0.3405$), and no main effect of stress ($F_{1,31} = 1.794$, $p = 0.1901$; Fig. 2B). When data were analyzed separately by sex (Fig. 2B insets), the results were similar but did not reach significance (for males: stress, $F_{1,14} = 0.2007$, $p = 0.6610$; extinction,

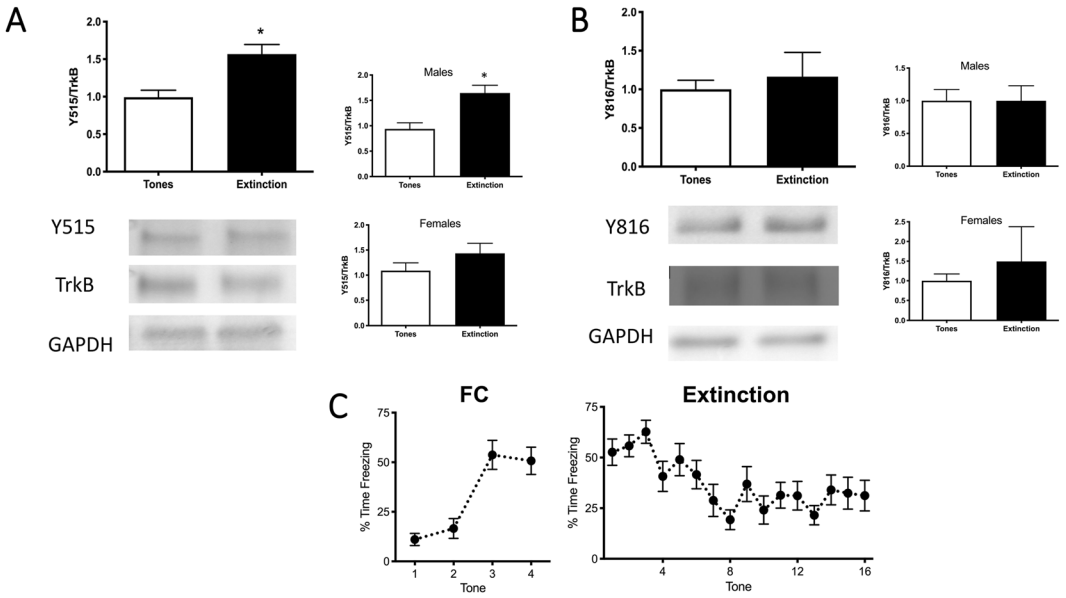


Fig. 1 Extinction induces the phosphorylation of the BDNF receptor TrkB at Y515, but not Y816 in the IL. A Phosphorylation of TrkB at Y515 is increased in the IL 30 min after the end of extinction compared to tone controls (*$p = 0.0018$, $n = 14$–18/group). Insets show data for males and females separately; a student's *t*-test showed an effect of extinction on Y515 in males ($t_{17} = 3.568$, $p = 0.0024$) but not in females alone. **B** Phosphorylation of TrkB at Y816 was unchanged 30 min after the end of extinction compared to tone controls ($p = 0.6292$, $n = 12$/group), insets show males and females separately. Bars represent SEM. **C** Fear conditioning (left) and extinction (right) curves showing % time freezing during the presentation of the tone. Bars represent SEM.

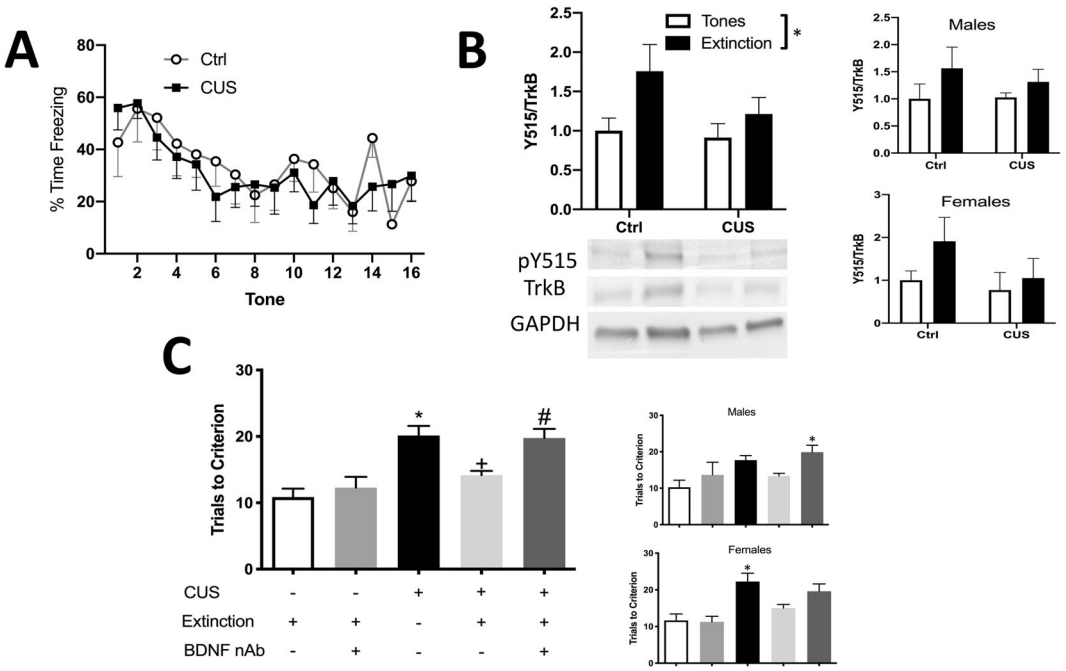


Fig. 2 Infralimbic BDNF is necessary for the effects of extinction on set shifting in stressed animals. A Fear extinction curves in stressed and non-stressed rats, showing percent time freezing during tone presentation. Bars represent SEM. **B** Extinction induced an increase in phosphorylation of TrkB at Y515 compared to tone controls (*$p = 0.0315$; 8–9 rats/group). Insets show males and females separately. **C** CUS impaired set shifting on the AST (CUS/Tones/IgG compared to No Stress/Extinction/IgG, *$p = 0.0003$), and extinction rescued set shifting in stressed rats (CUS/Extinction/IgG compared to CUS/Tones/IgG, +$p = 0.0091$). Blocking BDNF in the IL at the time of extinction attenuated the beneficial effect of extinction in stressed rats (CUS/Extinction/anti-BDNF compared to CUS/Extinction/IgG, #$p = 0.0218$). The BDNF antibody alone had no effect in non-stressed control rats (No stress/Extinction/anti-BDNF compared to No stress/Extinction/IgG, $p = 0.9739$). Insets show males and females separately. In males, multiple comparisons showed a difference between No stress/extinction/control IgG vs CUS/extinction/anti-BDNF. In females, multiple comparisons showed differences between No stress/extinction/control IgG vs CUS/tones/IgG. Bars represent SEM.

$F_{1,14} = 2.844$, $p = 0.1138$; interaction, $F_{1,14} = 0.3002$, $p = 0.5924$; for females: stress, $F_{1,13} = 1.794$, $p = 0.2034$; extinction ($F_{1,13} = 1.499$, $p = 0.2425$; interaction, $F_{1,13} = 0.5060$, $p = 0.4894$). There was no correlation between extinction, expressed as the freezing index

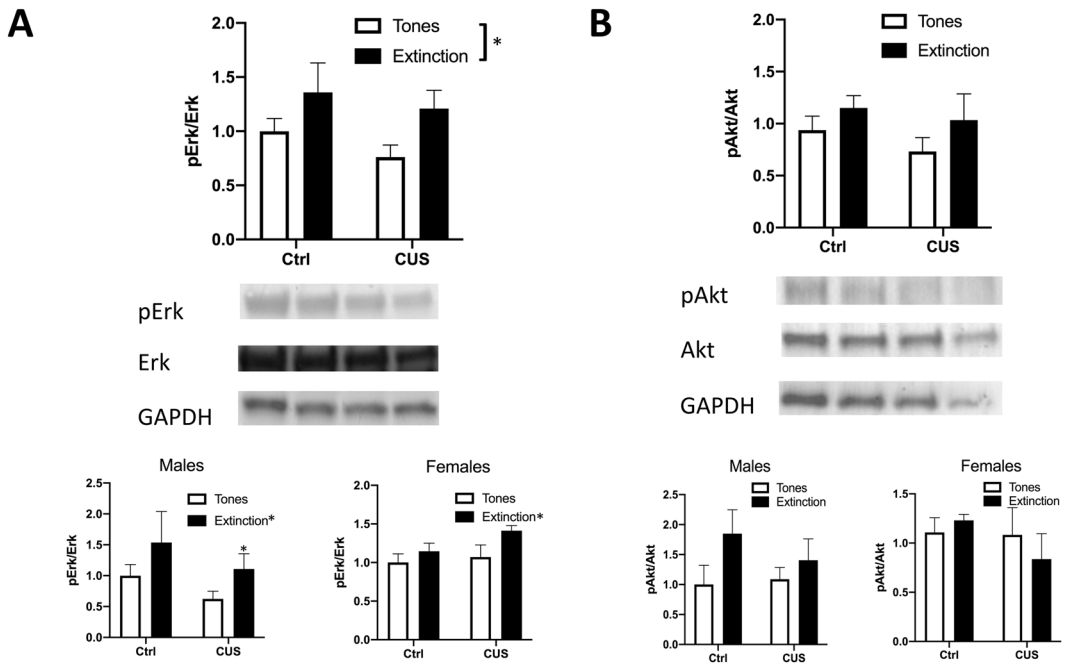


Fig. 3 Extinction induces the phosphorylation of Erk but not Akt in the IL. A Extinction increased phosphorylation of Erk in the IL 30 min after extinction compared to tone controls (*$p = 0.0021$, $n = 12$–14/group). Insets show males and females separately. ANOVA on male data showed a main effect of extinction on pErk ($F_{1,30} = 8.469$, $p = 0.0067$), a main effect of stress ($F_{1,30} = 5.136$, $p = 0.038$), and a difference between Tones CUS vs Extinction Control ($p = 0.0074$). Extinction induced an increase in pErk in females and a difference between Tones-CUS vs Extinction-CUS $p = 0.0023$). **B** Extinction did not significantly increase the phosphorylation of Akt in the IL 30 min after the end of extinction compared to tone controls ($p = 0.1639$, $n = 11$–14/group). Insets show males and females separately.

(mean freezing at plateau on tones 9–16 / mean initial freezing on tones 2–4) and pTrkB Y515 levels ($r_{18} = -0.2789$, $p = 0.2623$).

BDNF activity in IL cortex during extinction is necessary for its effects on set shifting

In total, 54 rats (28 males, 3–7/group; 26 females, 3–8/group) were used in five groups: (1) control group with control IgG in the IL (No stress/extinction/control IgG), (2) control group with neutralizing antibody against BDNF in the IL (No stress/extinction/anti-BDNF), (3) stress + tone control (CUS/tones/IgG), (4) extinction treatment (CUS/extinction/IgG), and 5) a group to test the necessity of BDNF for the effects of extinction (CUS/extinction/anti-BDNF). ANOVA revealed a significant group effect on set shifting ($F_{4,49} = 9.544$, $p < 0.0001$; Fig. 2C). Pairwise comparisons using the Holm–Sidak test showed that CUS impaired set shifting (CUS/Tones/IgG compared to No Stress/Extinction/IgG, $p = 0.0003$). Extinction rescued set shifting in stressed rats (CUS/Extinction/IgG compared to CUS/Tones/IgG, $p = 0.0091$), and blocking BDNF in the IL attenuated the beneficial effect of extinction (CUS/Extinction/anti-BDNF compared to CUS/Extinction/IgG, $p = 0.0218$). There was no effect of BDNF antibody alone on set shifting in non-stressed rats (No stress/Extinction/anti-BDNF compared to No stress/Extinction/IgG, $p = 0.9739$). These results suggest that BDNF activity in IL cortex is necessary during extinction for its beneficial effect on set shifting in stressed animals. As reported previously [17, 29] there was no effect of the treatments on fear extinction, as the extinction curves did not differ between groups ($F_{5,47} = 0.8359$, $p = 0.5310$). Analysis of male data alone revealed a significant group effect on set shifting ($F_{4,23} = 4.662$, $p = 0.0067$). Holm–Sidak multiple comparisons showed a difference between No stress/extinction/control IgG vs CUS/extinction/anti-BDNF ($p = 0.0110$); other group comparisons were not significant. ANOVA also revealed a significant group effect in females ($F_{4,21} = 5.726$, $p = 0.0028$); Holm–Sidak multiple comparisons showed differences between No stress/extinction/control IgG vs CUS/tones/IgG ($p = 0.0263$).

Extinction induces the phosphorylation of Erk but not Akt in the IL of stressed and control animals

In total, 51 rats (34 males, 8–10/group, 23 females, 4–7/group) were used in four groups (stress x extinction) to assess phosphorylation of Erk in the IL after extinction. ANOVA revealed a main effect of extinction ($F_{1,47} = 10.60$, $p = 0.0021$; Fig. 3A), no interaction ($F_{1,47} = 0.4968$, $p = 0.4844$), and no main effect of stress ($F_{1,47} = 2.375$, $p = 0.1300$). ANOVA on male data showed a main effect of extinction on pErk ($F_{1,30} = 8.469$, $p = 0.0067$), a main effect of stress ($F_{1,30} = 5.136$, $p = 0.038$), and no significant interaction ($F_{1,30} = 0.1545$, $p = 0.6971$). Holm–Sidak multiple comparisons showed a difference between Tones CUS vs Extinction Control ($p = 0.0074$). Extinction induced an increase in pErk in females ($F_{1,19} = 10.07$, $p = 0.005$) and a significant interaction ($F_{1,19} = 4.75$, $p = 0.0421$), but no main effect of stress ($F_{1,19} = 0.5888$, $p = 0.4523$). Holm–Sidak multiple comparisons showed a difference between Tones-CUS vs Extinction-CUS in females ($p = 0.0023$).

54 rats (25 males, 5–8/group; 18 females, 5–7/group) were used to assess pAkt. ANOVA revealed no significant main effect of stress ($F_{1,45} = 2.038$, $p = 0.1603$), no effect of extinction ($F_{1,45} = 0.8007$, $p = 0.3756$), and no interaction ($F_{1,45} = 0.05960$, $p = 0.8082$; Fig. 3B). Also, no significant effects were seen when data were analyzed separately by sex.

The PI3k-Akt and MAPK-Erk signaling pathways are involved in the therapeutic effects of extinction on set shifting after stress

In total, 62 rats (35 males, 3–7/group; 27 female rats, 3–5/group) were used in seven groups: (1) a control group (no stress/vehicle/extinction), (2) a control group with PI3k inhibitor LY294002 alone (no stress/LY294002/extinction), (3) a control group with MAPK inhibitor PD98059 alone (no stress/PD98059/extinction), (4) stress + tone control group (CUS/vehicle/tones), (5) extinction treatment (CUS/vehicle/extinction), (6) a group to determine the effect of blocking PI3k on the effects of extinction

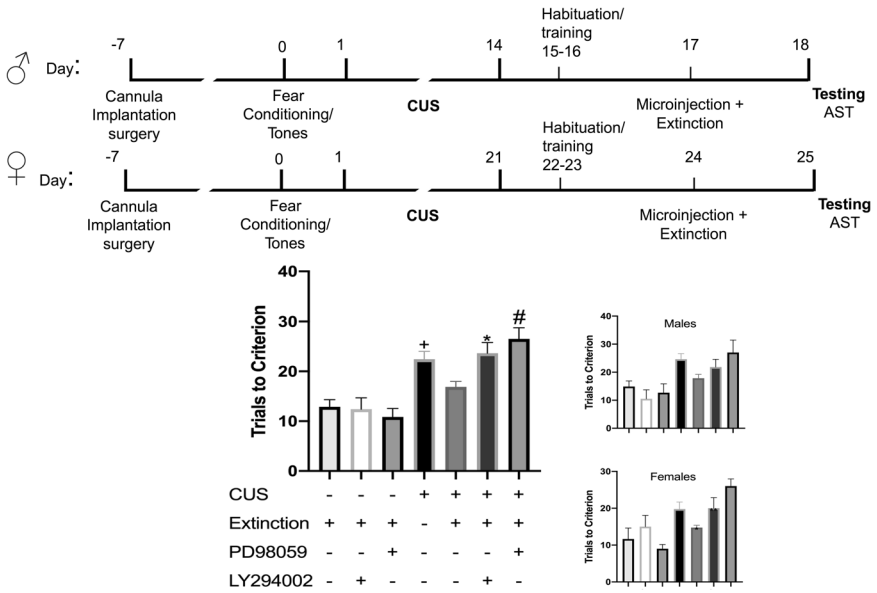


Fig. 4 Erk and Akt signaling are required in the IL cortex for the beneficial effects of extinction on set shifting in stressed animals. Microinjections of either LY200492 or PD98059 into the IL cortex immediately after extinction blocked the effects of extinction on set-shifting 24 h later in stressed rats. CUS induced a deficit in set shifting (CUS Tones Veh vs Ctrl Ext Veh, $+p = 0.0026$). Both PD98059 and LY294002 attenuated the improvement in set-shifting induced by extinction in stressed animals (CUS Ext Veh vs CUS Ext PD98059: $\#p = 0.0031$; CUS Ext Veh vs CUS Ext LY294002: $*p = 0.0389$). Insets show males and females separately.

(CUS/LY294002/extinction), and (7) a group to determine the effects of blocking MAPK on the effects of extinction (CUS/PD98059/extinction). ANOVA revealed a significant group effect ($p < 0.0001$, $F_{6,55} = 11.44$; Fig. 4). Holm–Sidak post hoc tests showed an effect of stress (CUS Tones Veh vs Ctrl Ext Veh, $p = 0.0026$). Neither inhibitor alone affected set shifting in control animals (Ctrl Ext Veh vs Ctrl Ext LY294002: $p = 0.95$, Ctrl Ext Veh vs Ctrl Ext PD98059: $p = 0.93$). Extinction again had a beneficial effect after stress, as the performance of extinction-treated stressed animals was comparable to control animals (Ctrl Ext Veh vs CUS Ext Veh, $p = 0.63$). Both PD98059 and LY294002 attenuated performance in stressed animals that received extinction (CUS Ext Veh vs CUS Ext PD98059: $p = 0.0031$; CUS Ext Veh vs CUS Ext LY294002: $p = 0.0389$). Therefore, blocking either MAPK/Erk or PI3k/Akt signaling immediately after extinction blocked the beneficial effects on set shifting tested 24 h after extinction in stressed animals. As above, there was no effect of the treatments on extinction itself, as the extinction curves did not differ between groups ($F_{3,34} = 0.5311$, $p = 0.6640$). One way ANOVA on male data only revealed a significant group effect ($F_{6,27} = 5.3879$, $p = 0.0009$, Fig. 4 inset). However, as this experiment was not specifically powered to analyze separately by sex, Holm–Sidak multiple comparisons revealed no significant differences between groups: CUS Ext Veh vs CUS Ext PD98059 ($p = 0.2896$), Ctrl Ext Veh vs CUS Tones Veh ($p = 0.0555$), CUS Ext Veh vs CUS Ext PD98059 ($p = 0.2896$), CUS Ext Veh vs CUS Ext LY294002 ($p = 0.7769$). One way ANOVA on female data also revealed a significant group effect ($F_{6,19} = 6.937981$, $p = 0.0005$), but Holm–Sidak multiple comparisons revealed no specific group differences: Ctrl Ext Veh vs CUS tones Veh ($p = 0.4167$), CUS Ext Veh vs CUS Tones Veh ($p = 0.7460$), CUS Ext Veh vs CUS Ext LY294002 ($p = 0.0546$) and CUS Ext Veh vs CUS Ext PD98059 ($p = 0.0662$).

BDNF alone reverses the effects of CUS on set shifting, and Erk signaling but not PI3k signaling is necessary for the effects of BDNF

In total, 24 rats (12 males, 3/group; 12 females, 3/group) were used in four groups: (1) a stress group (CUS/saline), (2) a group to determine whether BDNF alone microinjected into IL cortex in place of extinction is sufficient to reverse the effects of stress (CUS/BDNF), (3) a group to test whether the effects of BDNF are blocked by the PI3k inhibitor (CUS/BDNF/LY294002), and (4) a group to test whether the effects of BDNF are blocked by the MAPK inhibitor (CUS/BDNF/PD98059). ANOVA revealed a significant group effect ($F_{3,20} = 5.602$, $p = 0.0059$, Fig. 5). Multiple comparisons Holm–Sidak test showed that CUS-BDNF animals performed better than CUS-vehicle animals ($p = 0.0349$), while CUS-BDNF-PD90859 treated rats did not perform better than CUS-Veh animals ($p = 0.9616$). By contrast, stressed rats that received BDNF and the PI3k inhibitor still performed better than CUS-Veh animals ($p = 0.0408$), and were not significantly different than CUS-BDNF alone ($p = 0.9616$). Further, rats treated with BDNF and the Erk inhibitor PD90859 performed worse than CUS-BDNF ($p = 0.0385$). Therefore, BDNF alone in the IL cortex reversed the effects of stress on set shifting, and this effect was blocked by inhibiting Erk signaling, but not PI3k. One way ANOVA on male data alone revealed a significant group effect $F_{3,8} = 6.050$, $p = 0.0187$, Fig. 5). Holm–Sidak revealed no differences between CUS–Veh males compared to CUS-BDNF males ($p = 0.3273$), but did detect a difference between CUS-Veh and CUS/BDNF/LY294002 ($p = 0.0312$). No differences were observed in CUS-BDNF vs CUS/BDNF/PD98059 in males ($p = 0.3273$). A one way ANOVA revealed no group differences in female data analyzed separately ($F_{3,8} = 3.652$, $p = 0.0635$); thus, pairwise comparisons were not performed.

DISCUSSION

The results of this study demonstrate the necessity of BDNF activity in the IL for the beneficial effects of extinction on restoring stress-induced cognitive deficits in both male and female rats. Our findings are consistent with previous work highlighting the role of BDNF in extinction-mediated plasticity [21]. In humans, the BDNF Val66Met polymorphism, a single nucleotide polymorphism associated with reduced BDNF release, increases the risk of developing PTSD, and is associated with a poorer response to exposure therapy [33, 34]. Thus, compounds that enhance fear extinction by promoting BDNF activity may be clinically relevant to the treatment of PTSD and MDD [11]. Following exposure

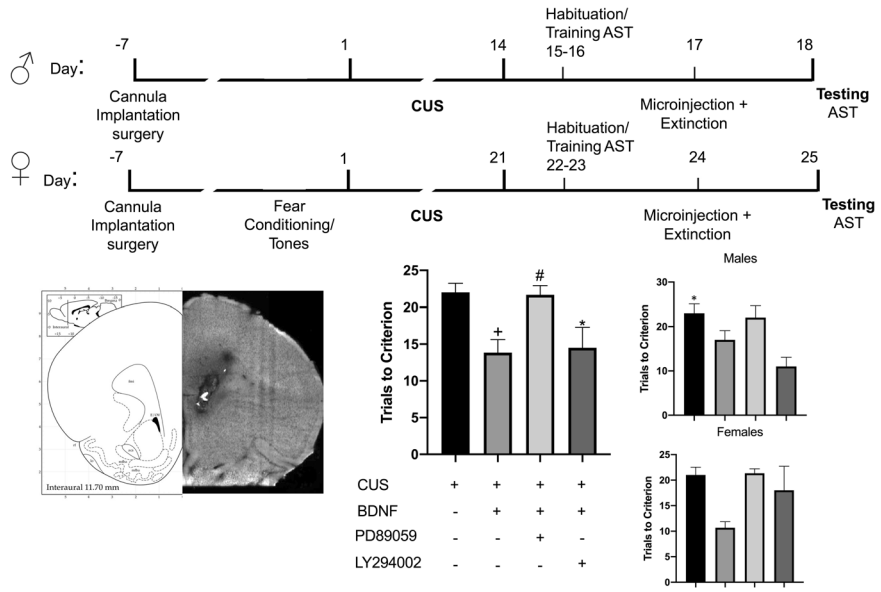


Fig. 5 Exogenous BDNF administered in the IL cortex reverses the effects of stress on set shifting, and these effects are dependent on Erk signaling. CUS-BDNF animals performed better than CUS-vehicle-treated animals ($+p = 0.0349$, $n = 6$ rats/group). Stressed rats that received BDNF and the PI3k inhibitor LY294002 performed better than CUS-Veh animals ($*p = 0.0408$), and were not significantly different than CUS-BDNF alone ($p = 0.9616$). By contrast, stressed rats treated with BDNF and the Erk inhibitor PD90859 performed worse than CUS-BDNF rats ($\#p = 0.0385$). Insets show males and females separately. In males, Holm–Sidak tests showed a difference between CUS-Veh and CUS/BDNF/LY294002 ($p = 0.0312$). Bars represent SEM. Bottom left is a representative example of injection site localization by dye injection into the IL.

therapy, symptoms may take weeks or months to improve, and a small number of patients achieve remission. Our results suggest that pharmacotherapies enhancing BDNF/Erk signaling in the mPFC may be advantageous as adjuncts to behavioral therapy. For example, studies have demonstrated the role of BDNF and Erk signaling in the antidepressant effects of ketamine [35], which is effective in treatment-resistant patients. Indeed, a recent pilot study showed that extinction and reconsolidation combined with adjunct administration of ketamine were effective in treating PTSD [36].

Several studies have demonstrated a role for BDNF in restoring mPFC function compromised by stress [37, 38]. Chronic stress induces functional and morphological alterations in the mPFC, such as dendritic atrophy of pyramidal neurons and reduced expression of glutamate receptors [39, 40], whereas BDNF induces long-term potentiation (LTP), enhances neuronal excitability, glutamate receptor trafficking, and promotes local protein synthesis in dendrites [41]. Extinction restores mPFC responsivity compromised by stress [17], and requires de novo protein synthesis in the IL for its beneficial effects. The current results suggest that BDNF activity in the IL during extinction facilitates the restoration of mPFC responsivity after stress, perhaps by promoting the synthesis of plasticity-related proteins. This hypothesis requires further testing. We observed phosphorylation of the BDNF receptor TrkB in IL after extinction, consistent with reports suggesting the activity-dependent release of BDNF during extinction [21]. After extinction learning, BDNF is increased in ventral hippocampal neurons [42]. Release of BDNF from ventral hippocampal projections to the IL is crucial for extinction memory [42]. Local release of BDNF elicits dendritic growth [43], a mechanism through which extinction-induced BDNF release in IL may reverse the detrimental effects of stress.

Inducing plasticity in the IL via extinction may also reverse stress-induced behavioral impairments mediated in other brain regions innervated by IL cortex. During extinction, the IL exerts inhibitory control over the amygdala to suppress fear. In individuals with PTSD, mPFC hypoactivity and amygdala hyperresponsivity are predictive of poor response to cognitive behavioral therapy [44]. Restoring IL function via BDNF signaling may therefore correct amygdala hyperresponsiveness and exaggerated fear response by enhancing activity in the mPFC. The IL also projects to the lateral septum (LS), which mediates active coping behavior in the shock probe defensive burying test (SPDB) [45, 46]. We have reported that extinction corrects the maladaptive shift from active to passive coping in the SPDB after chronic stress, and this effect relies on IL activity [17]. Thus, BDNF-induced plasticity in IL cortex during extinction may improve active coping behavior by facilitating modulatory regulation of LS activity by IL after stress.

The current results also elucidate intracellular signaling mechanisms potentially involved in the effects of extinction. Both Akt and Erk signaling contributed to the beneficial effects of extinction on set shifting that had been compromised by stress. Both Akt and Erk signaling contribute to long-term memory formation [47, 48]. Therefore, inhibiting these pathways likely interferes with extinction memory consolidation in the IL, which may reduce the subsequent therapeutic effects of extinction after stress. Extinction induced a significant increase in Erk phosphorylation, but only a modest, non-significant increase in phosphorylated Akt, consistent with previous reports [22]. This may seem at odds with the demonstration that PI3k blockade attenuated the beneficial effects of extinction. However, it is possible that only a small fraction of the total pool of Akt is activated by extinction, or that an increase in Akt phosphorylation only occurs in some cells after extinction, making it difficult to detect against a large unaffected background. Our results suggest, however, that even if the increase in phosphorylation is modest after extinction, Akt signaling is nonetheless necessary for the beneficial effects of extinction on set shifting. Another possibility is that, in addition to BDNF activity, extinction may also invoke other processes and transmitters that require basal PI3k-Akt activity in the IL, independent of induction by BDNF, to promote plasticity beneficial to set shifting. For example, Insulin Growth Factor-2 signals via PI3k-Akt, and has been shown to be involved in the consolidation of fear extinction, making it another

potential candidate for inducing plasticity through actions of PI3k-Akt [49].

Administration of exogenous BDNF into the IL in lieu of extinction reversed the effects of stress on set shifting, and this required Erk, but not Akt signaling. Although Erk and Akt share common targets, these pathways have different roles in some BDNF-related processes [47–49]. These results are in agreement with studies demonstrating the importance of Erk signaling for memory- and BDNF-induced plasticity [50–52]. Erk activation is necessary for BDNF-induced dendritic spine formation [50], and extinction increases dendritic spine density after acute stress [53]. Moreover, BDNF requires Erk signaling to phosphorylate cyclic AMP response element-binding protein (CREB) which mediates memory phase transitions [54]. BDNF/Erk-induced CREB activation drives the expression of genes regulating LTP and neuronal plasticity, e.g., activity-regulated cytoskeleton-associated protein (Arc) and Zif268 [55]. Thus, it is likely that extinction, via BDNF-Erk signaling, induces structural and functional plasticity in the IL necessary for memory consolidation, and for the beneficial effects of extinction on stress-compromised cognitive function such as set shifting.

In sum, the current results replicated our previous reports that extinction reversed stress-induced deficits in set shifting [17, 29] in both males and females, and demonstrated that these beneficial effects of extinction are dependent on BDNF activity in the IL during extinction. Extinction phosphorylated TrkB at the Y515 residue, but not Y816, suggesting a potential role for Y515-initiated intracellular signaling for the effects of extinction. Inhibiting either Akt or Erk in the IL after extinction prevented the therapeutic effects of extinction in stressed animals, and extinction increased phosphorylation of Erk, but not Akt, in the IL cortex. We further showed that exogenous BDNF microinjected into the IL was sufficient to reverse the detrimental effects of stress on set shifting, mimicking the beneficial effects of extinction. The beneficial effects of BDNF alone were mediated by Erk, but not Akt signaling. This information may be useful in developing strategies for adjunctive treatment during exposure therapy for PTSD to enhance or accelerate therapeutic efficacy and improve treatment outcomes.

REFERENCES

1. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet. 2007;370:851–8.
2. Rytwinski NK, Scur MD, Feeny NC, Youngstrom EA. The co-occurrence of major depressive disorder among individuals with posttraumatic stress disorder: a meta-analysis. J Trauma Stress. 2013;26:299–309.
3. Wiles N, Thomas L, Abel A, Ridgway N, Turner N, Campbell J, et al. Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. Lancet 2013;381:375–84.
4. de Kleine RA, Rothbaum BO, Van, Minnen A. Pharmacological enhancement of exposure-based treatment in PTSD: a qualitative review. Eur J Psychotraumatology. 2013;4:21626.
5. Ben-Zion Z, Fine NB, Keynan NJ, Admon R, Green N, Halevi M, et al. Cognitive flexibility predicts PTSD symptoms: observational and interventional studies. Front Psychiatry. 2018;9:477.
6. Peters AT, Jacobs RH, Crane NA, Ryan KA, Weisenbach SL, Ajilore O, et al. Domain-specific impairment in cognitive control among remitted youth with a history of major depression. Early intervention psychiatry 2017;11:383–92.
7. Birrell JM, Brown VJ. Medial frontal cortex mediates perceptual attentional set shifting in the rat. J Neurosci 2000;20:4320–4.
8. Radley J, Sisti H, Hao J, Rocher AB, McCall T, Hof P, et al. Chronic behavioral stress induces apical dendritic reorganization in pyramidal neurons of the medial prefrontal cortex. Neuroscience 2004;125:1–6.
9. Beck AT. Cognitive therapy and the emotional disorders: Penguin; 1979.
10. Walter KH, Palmieri PA, Gunstad J. More than symptom reduction: changes in executive function over the course of PTSD treatment. J Trauma Stress: Off Publ Int Soc Trauma Stress Stud 2010;23:292–5.
11. Andero R, Ressler KJ. Fear extinction and BDNF: translating animal models of PTSD to the clinic. Genes, Brain Behav 2012;11:503–12.
12. Foa EB, Rothbaum BO. Behavioural psychotherapy for post-traumatic stress disorder. Int Rev Psychiatry 1989;1:219–26.
13. Milad MR, Quirk GJ. Fear extinction as a model for translational neuroscience: ten years of progress. Annu Rev Psychol 2012;63:129–51.
14. Paredes D, Morilak DA. A rodent model of exposure therapy: the use of fear extinction as a therapeutic intervention for PTSD. Front Behav Neurosci 2019;13:46.
15. Hassien AM, Shue F, Bernier BE, Drew MR. A mouse model of stress-enhanced fear learning demonstrates extinction-sensitive and extinction-resistant effects of footshock stress. Behavioural Brain Res 2020;379:112391.
16. Sierra-Mercado D, Padilla-Coreano N, Quirk GJ. Dissociable roles of prelimbic and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. Neuropsychopharmacology 2011;36:529–38.
17. Fucich EA, Paredes D, Saunders MO, Morilak DA. Activity in the ventral medial prefrontal cortex is necessary for the therapeutic effects of extinction in rats. J Neurosci 2018;38:1408–17.
18. King AP, Block SR, Sripada RK, Rauch SA, Porter KE, Favorite TK, et al. A pilot study of mindfulness-based exposure therapy in OEF/OIF combat veterans with PTSD: altered medial frontal cortex and amygdala responses in social–emotional processing. Front Psychiatry. 2016;7:154.
19. Soliman F, Glatt CE, Bath KG, Levita L, Jones RM, Pattwell SS, et al. A genetic variant BDNF polymorphism alters extinction learning in both mouse and human. Science 2010;327:863–6.
20. Choi DC, Maguschak KA, Ye K, Jang S-W, Myers KM, Ressler KJ. Prelimbic cortical BDNF is required for memory of learned fear but not extinction or innate fear. Proc Natl Acad Sci 2010;107:2675–80.
21. Peters J, Dieppa-Perea LM, Melendez LM, Quirk GJ. Induction of fear extinction with hippocampal-infralimbic BDNF. Science 2010;328:1288–90.
22. Slouzkey I, Maroun M. PI3-kinase cascade has a differential role in acquisition and extinction of conditioned fear memory in juvenile and adult rats. Learn Mem 2016;23:723–31.
23. Hugues S, Chessel A, Lena I, Marsault R, Garcia R. Prefrontal infusion of PD098059 immediately after fear extinction training blocks extinction-associated prefrontal synaptic plasticity and decreases prefrontal ERK2 phosphorylation. Synapse 2006;60:280–7.
24. Alonso M, Vianna MR, Izquierdo I, Medina JH. Signaling mechanisms mediating BDNF modulation of memory formation in vivo in the hippocampus. Cell Mol Neurobiol 2002;22:663–74.
25. Duman RS, Voleti B. Signaling pathways underlying the pathophysiology and treatment of depression: novel mechanisms for rapid-acting agents. Trends Neurosci 2012;35:47–56.
26. Barfield ET, Gerber KJ, Zimmermann KS, Ressler KJ, Parsons RG, Gourley SL. Regulation of actions and habits by ventral hippocampal trkB and adolescent corticosteroid exposure. PLoS Biol 2017;15:e2003000.
27. Infralimbic BDNF is necessary for the therapeutic effects of extinction after chronic stress in male and female rats [database on the Internet]. Society for Neuroscience. 2019.
28. Bulin SE, Hohl KM, Paredes D, Silva JD, Morilak DA. Bidirectional optogenetically-induced plasticity of evoked responses in the rat medial prefrontal cortex can impair or enhance cognitive set-shifting. Eneuro. 2020;7: ENEURO.0363-19.2019.
29. Fucich EA, Paredes D, Morilak DA. Therapeutic effects of extinction learning as a model of exposure therapy in rats. Neuropsychopharmacology 2016;41:3092–102. https://doi.org/10.1038/npp.2016.127.
30. Girotti M, Silva JD, George CM, Morilak DA. Ciliary neurotrophic factor signaling in the rat orbitofrontal cortex ameliorates stress-induced deficits in reversal learning. Neuropharmacology 2019;160:107791.
31. Carreño FR, Walch JD, Dutta M, Nedungadi TP, Cunningham JT. BDNF-TrkB pathway mediates NMDA receptor NR2B subunit phosphorylation in the supraoptic nuclei following progressive dehydration. J Neuroendocrinol 2011;23:894.
32. Kim S, Shin J-K, Yoon HS, Kim J-H. Blockade of ERK phosphorylation in the nucleus accumbens inhibits the expression of cocaine-induced behavioral sensitization in rats. Korean J Physiol Pharmacol 2011;15:389–95.
33. Felmingham KL, Dobson-Stone C, Schofield PR, Quirk GJ, Bryant RA. The brain-derived neurotrophic factor Val66Met polymorphism predicts response to exposure therapy in posttraumatic stress disorder. Biol Psychiatry. 2013;73:1059–63.
34. Zhang L, Benedek D, Fullerton C, Forsten R, Naifeh J, Li X, et al. PTSD risk is associated with BDNF Val66Met and BDNF overexpression. Mol Psychiatry. 2014;19:8–10.
35. Duman RS, Li N, Liu R-J, Duric V, Aghajanian G. Signaling pathways underlying the rapid antidepressant actions of ketamine. Neuropharmacology 2012;62:35–41.

36. Pradhan B, Mitrev L, Moaddell R, Wainer IW. d-Serine is a potential biomarker for clinical response in treatment of post-traumatic stress disorder using (R, S)-ketamine infusion and TIMBER psychotherapy: a pilot study. Biochimica et Biophysica Acta (BBA)-Proteins Proteom 2018;1866:831–9.
37. Xu H, Wang J, Zhang K, Zhao M, Ellenbroek B, Shao F, et al. Effects of adolescent social stress and antidepressant treatment on cognitive inflexibility and Bdnf epigenetic modifications in the mPFC of adult mice. Psychoneuroendocrinology 2018;88:92–101.
38. Tornese P, Sala N, Bonini D, Bonifacino T, La Via L, Milanese M, et al. Chronic mild stress induces anhedonic behavior and changes in glutamate release, BDNF trafficking and dendrite morphology only in stress vulnerable rats. The rapid restorative action of ketamine. Neurobiol Stress. 2019;10:100160.
39. Liston C, Miller MM, Goldwater DS, Radley JJ, Rocher AB, Hof PR, et al. Stress-induced alterations in prefrontal cortical dendritic morphology predict selective impairments in perceptual attentional set-shifting. J Neurosci 2006;26:7870–4.
40. Jett JD, Bulin SE, Hatherall LC, McCartney CM, Morilak DA. Deficits in cognitive flexibility induced by chronic unpredictable stress are associated with impaired glutamate neurotransmission in the rat medial prefrontal cortex. Neuroscience 2017;346:284–97.
41. Leal G, Comprido D, Duarte CB. BDNF-induced local protein synthesis and synaptic plasticity. Neuropharmacology 2014;76:639–56.
42. Rosas-Vidal LE, Do-Monte FH, Sotres-Bayon F, Quirk GJ. Hippocampal–prefrontal BDNF and memory for fear extinction. Neuropsychopharmacology 2014;39:2161.
43. Horch HW, Katz LC. BDNF release from single cells elicits local dendritic growth in nearby neurons. Nat Neurosci 2002;5:1177–84.
44. Patel R, Spreng RN, Shin LM, Girard TA. Neurocircuitry models of posttraumatic stress disorder and beyond: a meta-analysis of functional neuroimaging studies. Neurosci Biobehav Rev 2012;36:2130–42.
45. Vertes RP. Differential projections of the infralimbic and prelimbic cortex in the rat. Synapse 2004;51:32–58.
46. Treit D, Pesold C, Rotzinger S. Dissociating the anti-fear effects of septal and amygdaloid lesions using two pharmacologically validated models of rat anxiety. Behav Neurosci 1993;107:770.
47. Slouzkey I, Rosenblum K, Maroun M. Memory of conditioned taste aversion is erased by inhibition of PI3K in the insular cortex. Neuropsychopharmacology 2013;38:1143–53.
48. Kelly Á, Laroche S, Davis S. Activation of mitogen-activated protein kinase/extracellular signal-regulated kinase in hippocampal circuitry is required for consolidation and reconsolidation of recognition memory. J Neurosci 2003;23:5354–60.
49. Agis-Balboa RC, Arcos-Diaz D, Wittnam J, Govindarajan N, Blom K, Burkhardt S, et al. A hippocampal insulin-growth factor 2 pathway regulates the extinction of fear memories. EMBO J 2011;30:4071–83.
50. Alonso M, Medina JH, Pozzo-Miller L. ERK1/2 activation is necessary for BDNF to increase dendritic spine density in hippocampal CA1 pyramidal neurons. Learn Mem 2004;11:172–8.
51. Mullen LM, Pak KK, Chavez E, Kondo K, Brand Y, Ryan AFRas/p38. and PI3K/Akt but not Mek/Erk signaling mediate BDNF-induced neurite formation on neonatal cochlear spiral ganglion explants. Brain Res 2012;1430:25–34.
52. Schafe GE, Atkins CM, Swank MW, Bauer EP, Sweatt JD, LeDoux JE. Activation of ERK/MAP kinase in the amygdala is required for memory consolidation of pavlovian fear conditioning. J Neurosci 2000;20:8177–87.
53. Moench KM, Maroun M, Kavushansky A, Wellman C. Alterations in neuronal morphology in infralimbic cortex predict resistance to fear extinction following acute stress. Neurobiol Stress. 2016;3:23–33.
54. Fukushima H, Zhang Y, Kida S. Active transition of fear memory phase from reconsolidation to extinction through ERK-mediated prevention of reconsolidation. J of Neurosci. 2021;41:1288–300.
55. Ying S-W, Futter M, Rosenblum K, Webber MJ, Hunt SP, Bliss TV, et al. Brain-derived neurotrophic factor induces long-term potentiation in intact adult hippocampus: requirement for ERK activation coupled to CREB and upregulation of Arc synthesis. J Neurosci 2002;22:1532–40.

ACKNOWLEDGEMENTS

We thank Christina George and Carlie McCartney for technical assistance.

AUTHOR CONTRIBUTIONS

D.P. contributed to data collection, analysis and interpretation, and wrote and edited the manuscript. A.K. contributed to data collection, interpretation, and edits to the manuscript. D.M. contributed to experimental design and conception, data interpretation, and editing as well as final approval of the manuscript.

FUNDING

This work was supported by research grant MH053851 from the National Institute of Mental Health, National Institutes of Health; by Merit Award 1I01BX003512 from the U.S. Department of Veterans Affairs Biomedical Laboratory Research and Development Program; and by a grant from the William and Ella Owens Medical Research Foundation, none of which had any role in study design, data collection, analysis or interpretation, nor in the preparation or decision to submit this paper for publication. The contents of this paper do not represent the views of the Department of Veterans Affairs or the U.S. Government.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to David A. Morilak.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021

Neuroscience and Biobehavioral Reviews 49 (2015) 19–31

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review

Inoculation stress hypothesis of environmental enrichment

Elizabeth J. Crofton, Yafang Zhang, Thomas A. Green*

Center for Addiction Research, Mitchell Center for Neurodegenerative Diseases, Department of Pharmacology and Toxicology, University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77555, United States

ARTICLE INFO

Article history:
Received 18 July 2014
Received in revised form 28 October 2014
Accepted 21 November 2014
Available online 29 November 2014

Keywords:
Environmental enrichment
Inoculation stress
Resilience
Drug addiction
Corticosterone

ABSTRACT

One hallmark of psychiatric conditions is the vast continuum of individual differences in susceptibility vs. resilience resulting from the interaction of genetic and environmental factors. The environmental enrichment paradigm is an animal model that is useful for studying a range of psychiatric conditions, including protective phenotypes in addiction and depression models. The major question is how environmental enrichment, a non-drug and non-surgical manipulation, can produce such robust individual differences in such a wide range of behaviors. This paper draws from a variety of published sources to outline a coherent hypothesis of inoculation stress as a factor producing the protective enrichment phenotypes. The basic tenet suggests that chronic mild stress from living in a complex environment and interacting non-aggressively with conspecifics can inoculate enriched rats against subsequent stressors and/or drugs of abuse. This paper reviews the enrichment phenotypes, mulls the fundamental nature of environmental enrichment vs. isolation, discusses the most appropriate control for environmental enrichment, and challenges the idea that cortisol/corticosterone equals stress. The intent of the inoculation stress hypothesis of environmental enrichment is to provide a scaffold with which to build testable hypotheses for the elucidation of the molecular mechanisms underlying these protective phenotypes and thus provide new therapeutic targets to treat psychiatric/neurological conditions.

© 2014 Elsevier Ltd. All rights reserved.

Contents

1. A history of environmental enrichment research 20
2. What is environmental enrichment? 20
   2.1. What are the beneficial effects of environmental enrichment? 20
   2.2. The protective addiction phenotype 21
   2.3. The protective antidepressant phenotype 21
   2.4. Anxiety 21
   2.5. Other disorders 21
   2.6. Species differences 21
3. What is inoculation stress? 22
4. Allostasis and allostatic load 22
5. Stress, addiction and depression 23
6. Does cortisol/corticosterone equal stress? 23
   6.1. CORT levels fluctuate throughout the day 24
   6.2. Rewarding stimuli induce CORT 24
   6.3. CORT induction is attenuated with chronic stress 24
   6.4. CORT is reinforcing 24
   6.5. Responses to CORT differ from responses to stress 24

* Corresponding author at: Center for Addiction Research, Department of Pharmacology and Toxicology, The University of Texas Medical Branch, 301 University Dr., Bldg. 17, 3.324G, Galveston, TX 77555, United States. Tel.: +1 409 747 7056; fax: +1 409 747 7050.
E-mail address: tom.green@utmb.edu (T.A. Green).

http://dx.doi.org/10.1016/j.neubiorev.2014.11.017
0149-7634/© 2014 Elsevier Ltd. All rights reserved.

7. Are enriched rats really stressed? 24
8. CORT as a possible mediator of the protective EC phenotype 26
9. Stress influences on self-administration 26
10. What is the best control for enrichment? 26
11. Conclusions 27
Acknowledgments 27
References 27

1. A history of environmental enrichment research

The "nature vs. nurture" debate began in earnest during the Victorian period, championed by Sir Francis Galton, who was inspired by the works of his cousin Charles Darwin. At issue was whether a person's expressed traits are a product of heritability (i.e. nature) or by his/her own experiences (nurture). Galton, bolstered by Darwin's theories on heritability came down firmly on the side of "nature". The opposing "nurture" side of the debate was best defined centuries before by John Locke's borrowed term "tabula rasa" (i.e. blank slate). The "nurture" side of the argument was further strengthened in the early 1900s by John Watson's theories on behaviorism.

As science evolved (particularly the advent of genetics), the "nature vs. nurture" debate evolved into a "genes vs. environment" debate, respectively. The battle raged on as scientists on both sides of the argument produced irrefutable evidence for their view. Eventually, scientists realized that both arguments were correct— that a person's expressed phenotype was due to an interaction of genes with environment. Thus, the Gene/Environment Interaction Theory was born. In a basic sense, the environment controls (to some degree) how genes are expressed. Thus, gene transcription is where the proverbial "rubber hits the road" and seems to play a significant role in the protective phenotypes produced by environmental enrichment (Green et al., 2010; Lobo et al., 2013; Zhang et al., 2014), which are described below in the beneficial effects of environmental enrichment.

The beginning of modern environmental enrichment research is mostly attributed to Rosenzweig, Renner, Bennett, Diamond and colleagues. This group used the environmental enrichment paradigm to show convincingly that the adult brain still exhibits plasticity and that, just like muscles, brains get stronger with greater use. Rats reared in an enriched condition (EC) have a thicker cortex, more dendritic arborization and greater cognitive abilities than rats reared in an isolated condition (IC) (Diamond et al., 1964; Renner and Rosenzweig, 1987; Rosenzweig and Bennett, 1996). Following these early experiments, many others have used environmental enrichment and found it to be a useful animal model in a variety of fields, particularly because it is a non-drug and non-surgical manipulation.

In parallel with Rosenzweig and colleagues, Harry Harlow was finalizing the ideas for his seminal work on the importance of maternal and social enrichment in rhesus monkeys (Harlow, 1958). Harlow designed inanimate wire and cloth "surrogate" mothers to show that maternal contact is enriching to baby macaques beyond merely providing food. Although Harlow's early work was oriented to the positive effects of maternal enrichment (i.e. affection), his later work shifted perspective to focus on the isolation aspect (i.e. lack of enrichment) rather than the enrichment itself (Harlow and Suomi, 1971).

2. What is environmental enrichment?

Environmental enrichment is complex and there are numerous ways to provide enrichment. There is a lack of consistency in protocols for enrichment between different laboratories, but the most common procedure in rats involves rearing the subjects in a large cage with novel objects and social contact with conspecifics for at least 30 days beginning immediately after weaning. The objects are replaced and rearranged daily to maximize novelty. This arrangement provides three key facets of enrichment: novelty, social contact and exercise. It has been shown in rats that all three aspects are rewarding (Bardo and Bevins, 2000; Belke, 2000; Bevins and Bardo, 1999) and all three release dopamine in the nucleus accumbens (Greenwood et al., 2011; Louilot et al., 1986; Rebec et al., 1997). Thus, it can be said that environmental enrichment is a compound manipulation that provides a daily workout for the dopamine system. Indeed, when the novel objects are replaced each day, the rats display a burst of exploratory activity lasting approximately 30 min that is beyond anything seen with locomotor stimulants like cocaine or amphetamine. Additionally, there is a second burst of exploratory/play behavior that occurs at the onset of the dark cycle, the beginning of the rats' normal period of high activity.

Although environmental conditions have a dramatic impact on the behavior of animals, these differing protocols for enriching rats often produce conflicting results. Parameters such as age of the animal, degree of enrichment, duration of enrichment, species and sex can each affect the results of an experiment. The lack of consistency in protocols likely stems from a lack of consensus regarding the definition of what indeed constitutes "environmental enrichment". Some might define enrichment based on environmental complexity—that a more complex environment is more enriching; however, environmental complexity alone is not the whole story. Environmental enrichment, by most definitions, should exert a positive influence on the organism, setting enrichment apart from overtly stressful events that have a negative impact on the organism. Thus, enrichment must provide an overall benefit to the organism. Further confusion in the field also arises from the fact that some researchers compare EC rats only to pair-housed social condition (SC) rats or compare only IC with SC rats (see below for discussion of the appropriate control for enrichment). However, without discounting or dismissing the views of others studying environmental enrichment using different protocols, this paper outlines a theory that the mild daily stresses of the enriched lifestyle are adaptive and inoculate rats to produce protective preclinical phenotypes for addiction and depression.

2.1. What are the beneficial effects of environmental enrichment?

As mentioned above, environmental enrichment contains three basic components: novelty, exercise and social contact. Animals are group-housed in a large cage equipped with children's plastic toys, which are replaced and rearranged every day. In order to study the "preventive" effect of environmental enrichment, rats are usually raised in the enriched condition before exposure to drugs (in the case of addiction research) or stress (in the case of depression research). Environmental enrichment attenuates the reinforcing effects of addictive drugs and produces an antidepressant-like effect (Brenes et al., 2008; Brenes Saenz et al., 2006; Green et al., 2010; Laviola et al., 2008). In addition, environmental enrichment can be studied as a "treatment" model, in which rats are assigned

to either an isolated or enriched condition *after* they are exposed to drugs or stress, which has also been shown to produce adaptive consequences (Grimm et al., 2008; Solinas et al., 2008; Thiel et al., 2010). Below is an overview of the beneficial effects of environmental enrichment. To maintain focus, this hypothesis paper is predominantly centered on rodent research, although environmental enrichment has been studied in other species with success (Harlow, 1958; Harlow and Suomi, 1971; Nader et al., 2012; Solinas et al., 2010).

2.2. The protective addiction phenotype

Rats reared in an enriched condition exhibit lower basal locomotor activity than rats in the isolated condition, making interpretation of drug-stimulated locomotor data challenging. Despite this, the available evidence is fairly clear that EC rats show greater locomotor sensitivity to psychostimulants such as amphetamine and cocaine (Bardo et al., 1999; Bowling and Bardo, 1994a; Bowling et al., 1993; Smith et al., 2009) while at the same time showing reduced sensitization to repeated exposure (Bardo et al., 1995; Green et al., 2003; Smith et al., 1997). In the conditioned place preference (CPP) paradigm, a paradigm more relevant to addiction, Bowling and Bardo, and then Green and colleagues reported that enrichment produces enhanced CPP to both amphetamine and cocaine in rats (Bowling and Bardo, 1994b; Green et al., 2010). These results lead to the hypothesis that enrichment *increases* the risk for addiction; however, in the paradigm with the greatest face validity for addiction, the intravenous drug self-administration paradigm, the reinforcing effects of amphetamine, cocaine, ethanol, and methylphenidate are decreased by enrichment and the rats self-administer less of these drugs (Alvers et al., 2012; Bardo et al., 2001b; Deehan et al., 2007; Gill et al., 2013; Green et al., 2002, 2010; Stairs et al., 2006).

Enrichment also alters drug taking in treatment models. Exposure to enrichment *after* exposure to cocaine reduces locomotor activity, eliminates cocaine CPP and reduces cocaine-induced reinstatement of CPP, and decreases cocaine-seeking behavior during extinction and cue-elicited reinstatement (Chauvet et al., 2012; Ranaldi et al., 2011; Solinas et al., 2008; Thiel et al., 2009). Although cue-elicited reinstatement is reduced in the treatment model, enrichment does not alter cocaine-primed reinstatement, suggesting that enrichment reduces the salience of drug-associated environmental cues which could lead to an effective therapy for craving elicited by drug cues in humans (Thiel et al., 2009).

2.3. The protective antidepressant phenotype

Along with the protective addiction phenotype, environmental enrichment also produces a protective antidepressant-like phenotype. In humans, three of the hallmark symptoms of depression are anhedonia, social withdrawal and behavioral despair. Our prior research and others found that, compared to IC rats, EC rats consume more sucrose in a sucrose preference test, indicating decreased anhedonia-like behavior; longer grooming time in the social contact test, suggesting decreased social withdrawal; and greater mobility time in the forced swim test (FST), suggesting reduced "behavioral despair" (Brenes et al., 2008; Brenes Saenz et al., 2006; Green et al., 2010).

2.4. Anxiety

Similar to addiction- and depression-like behavior, multiple labs have demonstrated reduced anxiety-like behavior resulting from environmental enrichment. For example, EC rats display lower basal locomotor activity, yet increased distance traveled in the center of the arena in the open field test, indicating an anxiolytic effect (Urakawa et al., 2013). In addition, enriched rats and mice were found to spend more time in the open arms in the elevated plus maze (EPM), and showed lower amounts of defensive burying and less defensive behavior when in close proximity to a predator, also suggesting reduced anxiety (Friske and Gammie, 2005; Leal-Galicia et al., 2007; Roy et al., 2001). Enriched mice show reduced anxiety in response to social defeat stress, an effect that was abolished by lesioning the infralimbic region of the prefrontal cortex prior to environmental enrichment exposure (Lehmann and Herkenham, 2011), suggesting a role of the prefrontal cortex in anxiety and environmental enrichment.

However, not all are convinced that enrichment is anxiolytic. Our own research demonstrated greater sucrose neophobia in EC rats and more fecal boli in the cold-stress defecation test, both suggesting *increased* anxiety (Green et al., 2010). In addition, latency to ejaculation has been used as a measure of anxiety (Wallace et al., 2009), and EC rats exhibit increased latency to ejaculate, again suggesting increased anxiety (Urakawa et al., 2014). It is noteworthy that the last three tests (sucrose neophobia, cold-stress defecation, and latency to ejaculation) are not a function of exploratory activity in a novel environment. Because environmental enrichment involves extended exposure to novel environments whereas IC rats have very little to no experience in novel environments, this confounds the results of behavioral tests involving exploration of a novel environment. In addition, environmentally enriched rats show less social withdrawal than isolated rats, which may confound the results of social defeat tests on enriched animals (Green et al., 2010). Taken as a whole, the benefits of enrichment on anxiety-like behavior are not as clear as other areas, but further analysis of anxiety-like behavior in enriched animals is an interesting avenue for future experimentation.

2.5. Other disorders

Environmental enrichment has also been shown to have beneficial effects in rodent models of neurodegenerative diseases. For example, in a mouse model of Huntington's disease, enrichment delays the onset and slows disease progression by minimizing the loss of cerebral volume and by rescuing protein deficits (Hockly et al., 2002; Spires et al., 2004; van Dellen et al., 2000). Enrichment can also prevent neurodegeneration in C57BL/6 mice caused by a neurotoxin that causes Parkinson's disease-like symptoms in humans by regulating expression of the dopamine transporter (DAT) (Bezard et al., 2003). In addition, an increasing number of studies reported the beneficial effect of environmental enrichment in improving learning and memory in behavioral and molecular aspects of Alzheimer's disease (Bouet et al., 2011; Jankowsky et al., 2003, 2005; Levi et al., 2003; Wolf et al., 2006). A study on Tg2576 mice, a model of Alzheimer's, found that environmental enrichment counteracts the deleterious effects of chronic unpredictable stress in Alzheimer's disease progression (Jeong et al., 2011). Further, a recent proteomic study found that the microtubule-associated protein tau was upregulated in EC rats compared to IC rats (Fan et al., 2013a,b). That paper and two other proteomics papers also identified huntingtin, presenilin 1, tau and amyloid precursor protein as major upstream regulators for environmental enrichment (Fan et al., 2013a,b; Lichti et al., 2014). These results warrant further investigation of neurodegenerative disorders using the environmental enrichment paradigm.

2.6. Species differences

Although environmental enrichment has been studied extensively in rats, other species have garnered considerably less attention, save for the work of Harlow in monkeys described above (Harlow, 1958; Harlow and Suomi, 1971). However, the available

evidence from the primate literature suggests that enrichment is a protective factor for stress, as well as for addiction-related behavior (Harlow, 1958; Harlow and Suomi, 1971; Kozorovitskiy et al., 2005; Nader et al., 2012). There are, however, some important rat/mouse differences in the effects of environmental enrichment. For example, rats show the unusual phenotype where environmental enrichment increases responsiveness to stimulants such as cocaine or amphetamine in locomotor, CPP and neurochemical studies, while these same rats show decreased drug self-administration (Bardo et al., 1999, 2001a; Bowling and Bardo, 1994a; Bowling et al., 1993; Green et al., 2002, 2010); mice, on the other hand, show decreased sensitivity to stimulants after environmental enrichment (Solinas et al., 2009). Regardless, environmental enrichment produces a net benefit in addiction related behavior in both species (Solinas et al., 2010).

3. What is inoculation stress?

As described above, the environmental enrichment paradigm is a non-drug, non-surgical preclinical animal model useful for studying various psychiatric and neurological conditions. Environmental enrichment produces protective phenotypes in addiction and depression models, which are robust and replicable. The major question is *how* environmental enrichment can produce such robust individual differences in a wide range of behaviors associated with addiction, depression and anxiety and beneficial effects even for animal models of neurodegenerative diseases. Our hypothesis is that enriched rats undergo inoculation stress. In short, chronic very mild stress from living in a complex environment and interacting non-aggressively with conspecifics inoculates enriched rats against subsequent stressors and/or drugs of abuse.

Inoculation stress, described previously in human studies, is a process of developing resilience to future stressful events by first being exposed to mildly stressful experiences early in life (Dienstbier, 1989; Fox et al., 2006; Khoshaba and Maddi, 1999; Lyons et al., 2009; Meichenbaum, 2007; Parker et al., 2004; Rutter, 2006). Exposure to stress or adversity that toughens an individual is protective, much like a vaccination that exposes an individual to a non-harmful version of a disease in order to develop immunity to that illness for the future (Lyons et al., 2009; Rutter, 2006). For example, adults who are exposed to work stress as adolescents have fewer negative health effects from work-related stress as adults (Mortimer and Staff, 2004). Unlike a vaccine, however, inoculation stress does not protect only against a single disease, but toughens the individual in general and prepares them to cope with a variety of stressors later in life. For example, individuals that have successfully coped with some adverse events in adolescence have overall better mental health and are better able to cope with serious illness, spousal loss, or a major accident as adults (Khoshaba and Maddi, 1999; Lyons et al., 2009; Seery et al., 2013).

An important distinction to make is that inoculation stress is not simply exposure to any stress early in life, but rather having positive and adaptive responses to mild stressors is critical. Severe stress early in life often causes the individual to be more vulnerable to stress later in life whereas mild stress exposure with an adaptive response can protect the individual (Khoshaba and Maddi, 1999; Parker and Maestripieri, 2011). Stress that is inoculating conditions the individual, and provides specific coping strategies for exposure to future stressors, ultimately leading to resilience.

Stress inoculation has also been described in non-human primates and rodents. Squirrel monkeys and rodents exposed to a manageable stress early in life are less stress-reactive in the future (Lyons et al., 2009, 2010). In particular, environmental enrichment in adolescence is able to provide protection from the deleterious effects of subsequent stressors (Fox et al., 2006; Larsson et al., 2002; Lyons et al., 2009; Parker et al., 2004), providing the impetus for the hypothesis that environmental enrichment is a chronic mild stress environment leading to stress inoculation.

4. Allostasis and allostatic load

In trying to make sense of how mild/moderate stressors are adaptive yet severe stressors are maladaptive, it helps to frame the picture with respect to allostasis and allostatic load. For the purposes of this paper, the term “allostasis” will be used in its broadest sense, disregarding the narrow application of this term only to energy regulation (McEwen and Wingfield, 2010). Thus, allostasis is defined as the process of returning a dynamic system to its stable set point after a challenge to that system. In a conceptual sense, homeostasis is maintaining stability in a non-dynamic system while allostasis is regulating stability of a dynamic system.

The genome is chock full of allostatic mechanisms for maintaining stability in hundreds of dynamic systems. However, maintaining stability of a dynamic system is not without cost to the organism. This cost to the organism is termed “allostatic load” and can be paid in a currency as diverse as energy usage, ion concentrations (e.g. neuronal activity), or protein turnover. Regulatory systems have evolved to deal with allostatic load to a certain extent, but most systems are susceptible to “allostatic overload” if the demands on the system outpace the allostatic capacity. At that point, the system fails to maintain stability and pathology develops. In the arena of psychiatric conditions, allostatic overload is evident when severe stressors surpass the allostatic capacity of the person and induce pathological conditions such as post-traumatic stress disorder, major depression, addiction, and anxiety disorders (among others).

One thing that differentiates allostasis from homeostasis is that allostasis, by virtue of regulating a dynamic system, can *predict* future allostatic load and adjust its capacity accordingly in anticipation of that future load. For example, forcing the average person to run 12 miles will result in allostatic overload of many systems (energy, oxygen, joints, muscles) whereas the seasoned runner’s body has predicted the possibility of a 12 mile run and adjusted the allostatic capacity of these systems to match the high load. As a result, instead of being pathogenic, subsequent stressors can even be motivating to the individual, allowing them to thrive in adverse conditions. Thus, individual responses to the same stressful event can be highly variable, and the resulting effect on the individual depends on that individual’s allostatic capacity to handle stress.

Allostatic overload of stress response systems can lead to psychiatric disorders, but how can inoculation stress or exposure to mild stress become protective against developing psychiatric conditions? Unfortunately stress has garnered a negative connotation, making experiments on stress with humans and animal models sometimes ambiguous. In 1976, Hans Selye defined stress as “the nonspecific response of the body to any demand made upon it” (Selye, 1976). The term “stress” does not inherently specify whether the stimulus is adaptive (i.e. inoculation stress) or maladaptive (i.e. allostatic overload), but even in Selye’s own work it was often assumed that “stress” meant maladaptive stress. However, whether a stressor is positive or negative chiefly depends on the *allostatic capacity* of the individual. Most individuals exposed to stress are resilient to psychiatric conditions but those that are susceptible have a lower allostatic capacity for stress. Inoculation stress causes the organism to predict future stress and increase the allostatic capacity of the individual accordingly, providing resilience to subsequent stressors that might otherwise produce psychiatric conditions. Thus, environmental enrichment is hypothesized to increase allostatic load capacity from the repeated exposure to very mild stresses, inoculating against subsequent stressors.

Since allostatic capacity for stress is associated with psychiatric disorders, it is useful to assess allostatic load capacity for stress in rodents. There is a wide range of stressors from which to choose to preclinically examine allostatic load capacity. Mildly stressful stimuli include brief periods of restraint stress, mild footshock, brief exposure to temperature changes as in the cold stress-induced defecation paradigm, and handling by an experimenter (Gärtner et al., 1980; Green et al., 2010; Gregus et al., 2005; Rabasa et al., 2011b). Some severe stressors are long periods of immersion in cold water, long periods of food or water deprivation, and intense footshock (Rabasa et al., 2011b; Willner, 1997). One very severe stressor is social defeat by an aggressive conspecific (Covington and Miczek, 2005; Koolhaas et al., 1996; Müller et al., 2000). It is important to note that individual stressors can be considered mild or severe depending on duration and intensity. For example, brief maternal separation in rats is adaptive whereas lengthy maternal separation leads to allostatic overload and susceptibility in models of psychiatric conditions (Lyons et al., 2009; Sih, 2011). Thus, footshock, maternal separation, physical restraint, and temperature changes can be considered mild stressors or severe stressors.

Predictability also seems important; the so-called chronic unpredictable mild stress paradigm (CMS), despite the label of "mild", can be a strong enough insult to induce allostatic overload and prevent inoculation (see below). Therefore, evaluating the response of the rodent to various stressors can aid in gauging the allostatic capacity of the rodent (i.e. whether they are inoculated against future stressors).

5. Stress, addiction and depression

As mentioned in the previous section, severe stress can lead to various psychiatric disorders; therefore, understanding the consequences of adaptive and maladaptive stress is translationally relevant. Stress is not only implicated in conditions such as posttraumatic stress disorder, major depression, and anxiety disorders, but also drug use disorders. Severe stress in adolescence (such as being abused as a child) is associated with higher risk for alcoholism, substance abuse, depression, suicide attempts, obesity and poorer general health (Felitti et al., 1998; McEwen, 2000). Evidence suggests there is overlap in the underlying mechanisms for mental health disorders and drug use disorders (Levin et al., 2008; Worley et al., 2012), and psychological stress may be the common link. The environmental enrichment paradigm produces robust protective phenotypes for depression-like and drug abuse behaviors. Therefore, our hypothesis is not only does maladaptive stress increase the likelihood of psychological disorders and drug abuse, but also the protective effects of environmental enrichment may be due to adaptive responses to stress.

Regarding stress and depression, stress is the leading factor for both the development of and relapse to major depression (Hardeveld et al., 2013; Kendler et al., 1998, 1999; Morris et al., 2010). Stress in the workplace and major life events, such as the death of a spouse, can trigger depressive episodes and increase the risk of major depression (Heim et al., 1997; Tennant, 2001). Evidence suggests that alterations in the hypothalamo–pituitary–adrenocortical (HPA) axis, which is involved in coordinating the body's response to various stressors (i.e. managing allostatic load), are involved in depression. Cortisol is a glucocorticoid released by the adrenal cortex in response to environmental stimuli and individuals with higher cortisol reactivity to low stress conditions had more depressive symptoms over time than those with low cortisol reactivity (Morris et al., 2012). In order to study this phenomenon preclinically alterations in the hypothalamo–pituitary–adrenocortical axis should be evaluated.

Regarding addiction, stress is a leading factor contributing to relapse to drug use in humans (Pohorecky, 1991; Sinha, 2001; Sinha et al., 1999, 2006), an effect modeled by self-administration studies in rats (Ahmed and Koob, 1997; Erb et al., 1996). Stress exposure in humans, particularly to a severe stressor, significantly increases cocaine craving during abstinence (Pohorecky, 1991). In addition to contributing to relapse, there is increasing preclinical evidence that stress contributes to the initial development of addiction (Burke and Miczek, 2014; Covington and Miczek, 2005; Goeders, 2002; Piazza and Le Moal, 1998). For humans, stress increases the likelihood that someone will start smoking along with increasing the risk of relapse to cigarette smoking (Bruijnzeel, 2012). Additionally, exposure to stress increases cocaine craving (Sinha et al., 1999) and more stress-induced cocaine craving increases the likelihood of relapse in cocaine-dependent individuals (Sinha et al., 2006).

Beyond stress/addiction and stress/depression interactions, there is evidence linking addiction directly to depression (Levin et al., 2008; Worley et al., 2012). For example, individuals with major depression are more likely to smoke than the average person (Bruijnzeel, 2012). Major depression and substance use disorders are often comorbid in humans, and the symptoms are often more severe together than with only one disorder (Kessler et al., 2005; Pettinati et al., 2013). The rate of comorbidity of major depression with alcohol use disorders is 40.3%, and major depression with a drug use disorder is 17.2% (Pettinati et al., 2013). The high rate of comorbidity makes developing an effective treatment very complicated. If the mood disorder is solely substance-induced, controlling the drug use would solve both disorders, but if the mood disorder has some other etiology, then antidepressants may be required (Pettinati et al., 2013). Often treatment for individuals with these comorbid disorders is focused on one disorder or the other and disregards the fact that stress may underlie both disorders.

Depression and addiction phenotypes often go hand in hand not only in humans but also in rodents (Green et al., 2006, 2010; Green et al., 2008; Pettinati et al., 2013). For example, overexpression of a dominant negative inhibitor of CREB (inducible cAMP early repressor/ICER or a mutant CREB/mCREB) or knockdown of CREB in the NAc produces an antidepressant-like phenotype and also decreases cocaine self-administration in a similar manner to environmental enrichment (Green et al., 2006, 2010). Similarly, the transcription factor ΔFosB in the NAc is associated with stress and also cocaine-taking behavior (Vialou et al., 2010; Zhang et al., 2014). Thus, it is very difficult to tease apart addiction and depression phenotypes in humans and in rodent models. Therefore analysis of an animal model that addresses both disorders simultaneously, such as environmental enrichment, is very valuable. Further, the close ties between depression and addiction are hypothesized to explain how inoculation to stress can affect drug-taking behavior.

As a whole, the evidence linking stress, depression and addiction provides a plausible rationale for how the repeated mild stress of environmental enrichment can protect against addiction-related behavior.

6. Does cortisol/corticosterone equal stress?

Because stress is linked with psychological disorders and drug dependence disorders, assessing stress in preclinical models is translationally relevant but can prove to be difficult without a good objective measure of stress in animals. Often, measurements of plasma cortisol, or the rodent equivalent corticosterone (CORT), have been used as a measure of stress in a variety of experiments. CORT is the primary glucocorticoid released by the adrenal cortex as a final product of the HPA axis. The HPA axis is activated in response to environmental stressors and activation of this system releases CORT. As a result, many scientists use CORT as a de facto

indication of a subject's stress level. However, it is important to note that CORT does not equal stress. The following paragraphs will argue: (1) CORT levels fluctuate throughout the day independent of stress, (2) rewarding stimuli induce CORT release, (3) CORT induction is *attenuated* with chronic stress, (4) CORT itself is reinforcing and (5) behavioral responses to CORT administration alone do not mimic responses to stress. In addition, there are several caveats when attempting to extrapolate emotional and behavioral state from plasma CORT levels, not the least of which is that the mere act of acquiring a blood sample to measure CORT can be stressful itself, especially with high frequency sampling (Abelson et al., 2005). Therefore, evaluating stress in environmentally enriched animals based on CORT levels has contributed to differing hypotheses on whether enrichment or isolation is inherently stressful because of inconsistent findings of corticosterone levels between enriched and isolated animals (Konkle et al., 2010). However, this is not to say that CORT is not important, or that corticosterone is not involved in the beneficial effects of environmental enrichment. In fact, CORT and the HPA axis may be involved in the inoculation stress of environmental enrichment (see Section 8 below for elaboration). However, measured CORT levels do not provide a complete picture of the adaptive or maladaptive nature of the stress responses of an animal.

6.1. CORT levels fluctuate throughout the day

CORT measurements taken at different times during the circadian cycle will vary because CORT has a characteristic circadian rhythm. Regardless of stress level, CORT generally peaks as rats awaken just prior to the dark cycle and is lowest at the beginning of the light cycle (Allen and Kendall, 1967; Butte et al., 1976). The dark cycle is the period where the animals are awake and highly active versus the light cycle when the animals are mostly sleeping. Thus, if one were to use CORT as a de facto measure of stress, the circadian rhythm of CORT will likely produce confounding results depending on when blood is collected.

6.2. Rewarding stimuli induce CORT

Phasic CORT release subsequent to an environmental stimulus is generally assumed to be an indication that the animal is in a negative emotional state and that the stimulus had a negative impact. For example, social defeat, where a test subject is physically dominated by a more aggressive conspecific, is a noxious stressor and causes release of CORT in the defeated male (Buwalda et al., 2012). However, assumptions of a negative state are not always true. For example, sexual activity, a positive and rewarding stimulus, releases similar amounts of CORT as social defeat (Buwalda et al., 2012). Another stimulus, exercise, which is regarded as positive and rewarding, can also cause an increase in cortisol in humans (Buono et al., 1986; Deinzer et al., 1997; Rojas Vega et al., 2006). Voluntary exercise can also increase circulating CORT in Sprague–Dawley rats (Fediuc et al., 2006). In addition, rewarding drugs, including cocaine, cause CORT release (Moldow and Fischman, 1987; Torres and Rivier, 1992). Both rewarding and noxious stimuli cause alterations in CORT; therefore, CORT levels alone cannot differentiate between negative and positive stimuli.

6.3. CORT induction is attenuated with chronic stress

If CORT is a de facto measure of stress, one would expect repeated stress to increase CORT. Multiple studies have shown, however, that induction of CORT after a mild/moderate stressor attenuates with repeated exposure to the stressor in rats (Barnum et al., 2007; Carter et al., 2004; De Boer et al., 1990; Magarinos and McEwen, 1995; Mizoguchi et al., 2001; Natelson et al., 1988; Rabasa et al., 2011a). In humans, cortisol release also habituates with repeated exposures to the same stressor although there are individual differences in cortisol responses to stress (Deinzer et al., 1997; Gerra et al., 2001). This evidence suggests that repeated stress increases the allostatic capacity to future stressors, which is a possible mechanism for the inoculation stress underlying environmental enrichment effects.

6.4. CORT is reinforcing

Additional evidence that CORT is not the same as maladaptive stress is demonstrated by the fact that CORT itself has reinforcing properties. Rats will intravenously (Piazza et al., 1991) and orally self-administer CORT (Deroche et al., 1993), causing release of dopamine in the nucleus accumbens (NAc) (Graf et al., 2013). CORT injections also potentiate amphetamine self-administration at medium and high doses of amphetamine (10 and 30 μg/infusion) (Piazza et al., 1991). These experiments and others show that CORT is reinforcing at circulating levels similar to that released by mild stress (Piazza et al., 1991).

6.5. Responses to CORT differ from responses to stress

If CORT equals stress, then CORT administration should produce the same responses as stress. MacDougall and Howland found that rats injected with CORT versus rats exposed to 30 min of restraint stress (a mild stressor) had the same amount of circulating CORT, but only restrained rats showed changes in short- and long-term synaptic plasticity in the subiculum (MacDougall and Howland, 2013). Retana-Marquez et al. (1998) found that CORT injections were not able to mimic the behavioral effects of social defeat stress even at very high circulating plasma levels. Social defeat causes decreases in male sexual behavior and decreases in testosterone, whereas CORT injections do not (Retana-Marquez et al., 1998). Conversely, rats restrained for 6 h a day for 21 days did not show an increase in depression-like behavior in the forced swim test but rats injected with CORT did show an increase in depression-like behavior, suggesting that in some cases, elevated CORT can cause more maladaptive changes than mild restraint stress (Gregus et al., 2005). Thus, CORT and stress sometimes produce different behavioral effects, and CORT administration alone cannot reproduce the behavioral effects of stress; therefore CORT does not equal stress.

All told, it is clear that although stress usually releases CORT, the circulating level of CORT is not a direct measurement of stress level. Further, it is important to remember that animals in a chronically mild-stress environment show attenuated rather than potentiated CORT induction to stress or drugs. If CORT is not an adequate measure of stress, how can we determine if enriched rats are actually more stressed than isolated rats?

7. Are enriched rats really stressed?

The inoculation stress hypothesis of environmental enrichment proposes that enriched rats are repeatedly stressed. However, at first sight, it is exceedingly difficult to make this case. Young male Sprague–Dawley rats (unlike mice) typically establish dominance hierarchies through play behavior and, as long as there are no female rats in the vicinity, typically do not feel the need to challenge these hierarchies over time. Thus, in this rat enrichment utopia, fighting is rare, food is plentiful, space is expansive, and rats get all of the novelty, social contact (rats are social creatures), and exercise they desire. Additionally, it has repeatedly been proposed that enrichment is the "functional opposite of stress" (Fox et al., 2006; Solinas et al., 2010; Wright and Conrad, 2008). If true, how can one make the case that enriched rats are chronically stressed?

Table 1
Parallels between repeated stress and enrichment.

	Effect	Repeated mild stress	Environmental enrichment
Endocrine	Blunted CORT induction	Natelson et al. (1988), De Boer et al. (1990), Magarinos and McEwen (1995), Barnum et al. (2007), Rabasa et al. (2011b), Carter et al. (2004)	Stairs et al. (2011), Stairs and Bardo (2009), Skwara et al. (2012)
	Blunted ACTH induction	Gadek-Michalska and Bugajski (2003)	Skwara et al. (2012), Belz et al. (2003)
	Enlarged adrenals	Swanson and van de Poll (1983), Marti et al. (1993)	Mlynarik et al. (2004), Moncek et al. (2004)
	Blunted adrenaline release	Dobrakovova et al. (1990)	Moncek et al. (2004)
Neurobiology	ΔFosB accumulation in NAcc	Perrotti et al. (2004), Lobo et al. (2013)	Zhang et al. (2014), Lobo et al. (2013), Solinas et al. (2009)
	Attenuated immediate-early gene induction in NAcc	Shoji and Mizoguchi (2010) (cFos), Green et al. (2008) (ATF3),	Zhang et al. (2014), manuscript in preparation
Behavior	Higher sensitivity to locomotor activating effects of amphetamine and cocaine	Deroche et al. (1992), Lepsch et al. (2005)	Bowling et al. (1993), Bowling and Bardo (1994a,b), Smith et al. (2009), Green et al. (2010)
	Decreased stimulant self-administration	Matthews et al. (1996), Moffett et al. (2006), *however, see:* Carroll and Meisch (1984), Piazza et al. (1990), Goeders and Guerin (1994), Shaham and Stewart (1994), Miczek and Mutschler (1996), Kosten et al. (2000)	Bardo et al. (2001a,b), Green et al. (2002, 2010), Stairs et al. (2006), Thiel et al. (2009), Alvers et al. (2012), Puhl et al. (2012)
	Increased defecation under stress conditions	Jorge et al. (2010)	Green et al. (2010)
	Decreased locomotor activity to novelty	Cruz et al. (2012)	Bowling et al. (1993), Green et al. (2003, 2010)
Physiology	Heart rate returns to baseline more quickly after stress	Carter et al. (2004), Chen and Herbert (1995)	Sharp et al. (2002)
	Lower body weight	Harris et al. (2004)	Pena et al. (2009)

Although there are multiple good lines of evidence suggesting enrichment produces the functional opposite of stress (Fox et al., 2006; Solinas et al., 2010; Wright and Conrad, 2008), none of the reports gives a possible explanation for how enrichment produces this effect. The inoculation stress hypothesis of environmental enrichment outlined here posits that enrichment is a chronic mildly stressful condition that induces neuronal and neuroendocrine plasticity leaving enriched rats more resistant (i.e. greater allostatic capacity) to overtly stressful stimuli. Environmental enrichment exposes animals to novelty, social contact, and exercise and multiple studies have found that these variables cause stress-like responses. Acute voluntary exercise induces the secretion of CORT (Fediuc et al., 2006) and exposure to novelty will induce secretion of both CORT and adrenocorticotropic hormone (ACTH) in rats (Hennessy et al., 1979; Ostrander et al., 2006; Piazza et al., 1991; Larsson et al., 2002). Finally, rats housed together with conspecifics had higher circulating CORT levels than isolated rats suggesting that social contact also causes a stress-like response (Raz, 2013). However, in the absence of a truly reliable and objective measure of stress, one must rely on circumstantial evidence comparing the effects of enrichment with the effects of repeated mild stress (see Table 1).

As summarized in Table 1, there are numerous parallels between the effects of environmental enrichment and repeated mild stress, adding strength to the idea that enriched rats are chronically stressed. Due to the importance of stress as a contributing factor to depression and addiction (see above), the effects of chronic stress on the body have been studied in depth. Results of these studies have produced a clear picture of the endocrine, neurobiological and behavioral sequelae of chronic stress in humans and in rat models. Environmentally enriched animals have also been assessed for the same endocrine, neurobiological and behavioral effects. There is much evidence that repeated mild stress blunts CORT induction in response to subsequent stressors (Barnum et al., 2007; Carter et al., 2004; De Boer et al., 1990; Magarinos and McEwen, 1995; Natelson et al., 1988; Rabasa et al., 2011b) and environmental enrichment also results in blunted CORT induction to stress (Skwara et al., 2012; Stairs and Bardo, 2009; Stairs et al., 2011). Repeated exposure to severe stressors (i.e. those producing allostatic overload), such as social defeat stress, does not show reductions in CORT induction to subsequent stress (Barnum et al., 2007). Repeated stress also blunts stress-induced adrenocorticotropic hormone (ACTH) induction (Gadek-Michalska and Bugajski, 2003) and stress-induced release of adrenaline (Dobrakovova et al., 1990). Environmental enrichment also blunts ACTH induction (Belz et al., 2003; Skwara et al., 2012) and reduces stress-induced release of adrenaline (Moncek et al., 2004). In addition, enlarged adrenal glands have been found in environmentally enriched animals (Mlynarik et al., 2004; Moncek et al., 2004) and in animals exposed to repeated mild stress (Marti et al., 1993; Swanson and van de Poll, 1983). Environmental enrichment and repeated mild stress also both produce lower body weights (Harris et al., 2004; Pena et al., 2009) and following stress, the animal's heart rate returns to baseline more quickly (Carter et al., 2004; Chen and Herbert, 1995; Sharp et al., 2002).

Not only do environmental enrichment and repeated mild stress show the same endocrine consequences, they also show concordant neurobiological consequences in the nucleus accumbens. Environmental enrichment reduces induction of immediate early genes (IEGs) (Zhang et al., 2014), as does repeated stress (Alibhai et al., 2007; Green et al., 2008; Shoji and Mizoguchi, 2010). In contrast to other IEGs, environmental enrichment causes accumulation of basal ΔFosB protein in the nucleus accumbens (Lobo et al., 2013; Solinas et al., 2009), which also occurs in repeatedly stressed animals (Lobo et al., 2013; Perrotti et al., 2004).

Environmental enrichment also produces robust effects on behavior that show similarities to the behavior of repeatedly stressed animals. For example, enriched animals show increased defecation to a mild stressor (novel cage under cold conditions) (Green et al., 2010), and rats that were alternately restrained for 1 h or placed on a platform surrounded by water for 1 h a day for 5 days also show increased defecation (Jorge et al., 2010). Environmentally enriched animals are more sensitive to the locomotor activating effects of amphetamine (Bowling and Bardo, 1994a; Bowling et al.,

1993) and cocaine (Smith et al., 2009). Repeatedly stressed animals are also more sensitive to the locomotor activating effects of amphetamine and cocaine (Deroche et al., 1992; Lepsch et al., 2005). Despite being more sensitive to locomotor stimulants, environmental enrichment decreases spontaneous locomotor activity in response to a novel environment (Bowling et al., 1993; Green et al., 2003, 2010). After mild stress, animals also show a similar decrease in locomotor activity when placed in a novel environment (Cruz et al., 2012).

As described above and illustrated in Table 1, environmental enrichment and repeated mild stress have matching endocrine, physiological, neurobiological, and behavioral effects. Although this is circumstantial evidence, it supports the argument that environmentally enriched animals are in a state of chronic mild stress and this mild stress in adolescence inoculates against future stressors.

The mild stress of environmental enrichment, however, is distinct from the stress of paradigms such as the chronic mild stress (CMS) paradigm or the chronic unpredictable stress paradigm in that environmental enrichment constitutes predictable stress that produces adaptive responses. The chronic mild stress paradigm typically involves stressors such as food deprivation, water deprivation, brief exposure to another subject, lights on during the dark cycle, periods of titling the cage by 30 degrees, and long periods of wet bedding material which occur randomly throughout the week for several weeks (Murison and Hansen, 2001; Willner et al., 1987, 1992). The CMS procedure can reduce sucrose and saccharine preference after several weeks of this unpredictable stress exposure (Hatcher et al., 1997; Willner et al., 1987), an anhedonic-like effect that can be reversed by several weeks of treatment with a tricyclic antidepressant (Willner et al., 1987). However, the CMS paradigm does not typically have other depression-like effects, and under some conditions this paradigm can actually increase sucrose consumption, suggesting inconsistencies in the paradigm (Murison and Hansen, 2001). An inoculation stress interpretation of these inconsistent data would posit that the unpredictable nature of the CMS can elevate the "mild" stress to a level that induces allostatic overload and that the inconsistencies in anhedonic-like behavior are a function of degree. The CMS by Murison and Hansen (2001) may not have induced allostatic overload, thus producing an inoculating effect whereas the other two reports induced a more severe stress (Hatcher et al., 1997; Murison and Hansen, 2001; Willner et al., 1987). Indeed, Hatcher and Hansen reported only finding an anhedonic saccharin response when the CMS paradigm included food deprivation.

8. CORT as a possible mediator of the protective EC phenotype

The sections above argue that CORT is not the same as stress; however, that is not to say that CORT is irrelevant. In fact, CORT responses may contribute substantially to the environmental enrichment protective phenotypes. For example, the environmental enrichment protective addiction phenotype fits well with what is known of the influences of CORT on stimulant self-administration. As mentioned above, CORT itself can be self-administered by rats (Deroche et al., 1993; Piazza et al., 1991), but there also exists evidence that CORT plays a significant role in stimulant self-administration. For example, higher CORT induction was associated with greater self-administration of low unit doses of cocaine regardless of whether the rats were stressed with contingent footshock, noncontingent footshock or no footshock (Goeders and Guerin, 1996b). Next, acquisition of cocaine self-administration can be completely blocked by bilateral adrenalectomy, partially reduced by pharmacological inhibition of corticosterone release by metyrapone in rats, and self-administration can be partially recovered by adding CORT to the drinking water (Goeders and Guerin, 1996a). Given that stimulants induce CORT release and blocking that release blocks self-administration, it is likely that the *amount* of CORT release from a stimulant determines the ability of low doses of that stimulant to engender or maintain self-administration. Blunted CORT release from stimulants in EC rats (Stairs and Bardo, 2009) could be the underlying molecular mechanism whereby inoculation stress produces the protective EC addiction phenotype. Further investigations are warranted to test this hypothesis.

9. Stress influences on self-administration

As discussed above, acute stress in humans is a major factor in relapse to addiction, so it is not surprising that acute stress in rodents produces reinstatement of cocaine seeking (Erb et al., 1996). However, the question at hand is how prior stress (i.e., not during or immediately before the session) affects subsequent stimulant self-administration. The logic in the above sections suggests that inoculation stress blunts CORT induction and that a blunted CORT response leads to less stimulant self-administration. Hence, one would predict that prior repeated mild stress (i.e. inoculation stress) would decrease drug self-administration. Two reports show that rats exposed to short-term maternal separation stress as pups later show reduced acquisition of cocaine self-administration at low unit doses (Matthews et al., 1996; Moffett et al., 2006). Despite these reports however, there are several reports showing that repeated stress *increases* stimulant self-administration (Carroll and Meisch, 1984; Goeders and Guerin, 1994; Kosten et al., 2000; Miczek and Mutschler, 1996; Piazza et al., 1990; Shaham and Stewart, 1994). Multiple factors could account for this discrepancy. First, the stressors in some of these papers are severe stressors that would exceed the allostatic capacity of the rats. Second, it is possible (even likely) that inoculation stress is most pronounced in very young animals (as with maternal separation and environmental enrichment). Third, it is possible that many other factors affect self-administration and that one or more of these factors are at play in some of these experiments. Regardless, if the inoculation stress hypothesis of environmental enrichment is correct, one would predict that environmental enrichment would decrease stimulant self-administration, which is undoubtedly the case (Alvers et al., 2012; Bardo et al., 2001b; Green et al., 2002, 2010; Puhl et al., 2012; Stairs et al., 2006; Thiel et al., 2009). In any case, the fact remains that EC rats have blunted CORT responses, blunted CORT responses are associated with less self-administration, and EC rats self-administer stimulants less readily than IC rats which supports the inoculation stress hypothesis.

10. What is the best control for enrichment?

An important problem in the enrichment field is the difficulty of being able to compare results between labs because of inconsistencies in enrichment protocols and the use of different control groups. Environmental enrichment is a compound manipulation whereby rats are chronically exposed to novelty, social contact, and exercise. The most rigorous scientific approach would be to study each aspect individually and then in combination. For example, it would be nice to know the relative contributions of social contact vs. object novelty vs. exercise, and if the combination of these factors is redundant, additive or synergistic. However, it is not possible to fully separate these aspects because social interaction is a form of novelty and greatly increases activity (i.e. exercise). In addition, this approach would entail studying eight different conditions, and we as scientists have an ethical obligation to reduce as much as possible the number of animals used in biomedical research (Council, 2011).

Additionally, as the number of conditions increases, there rapidly comes a point of diminishing returns where the cost (in money and time) of running an increasing number of conditions exceeds the small incremental benefit of the knowledge gained. Thus, the number of conditions must be limited. The fact that novelty, social contact and exercise all fall under the umbrella term "environmental enrichment" presupposes some commonality among the three constituents of enrichment. Indeed, each of these factors is rewarding to rats, and each releases dopamine in the nucleus accumbens, a critical brain region involved in stress, addiction, and depression (Greenwood et al., 2011; Louilot et al., 1986; Rebec et al., 1997). Accordingly, the many different conditions can now be reduced to as few as two: environmental enrichment and the appropriate control group.

Identifying the appropriate control group for cocaine administration is easy—an injection with no cocaine. Logic would dictate that the appropriate control for environmental enrichment (composed of novelty, social contact and exercise) would be a group with the absence of novelty, social contact, and exercise (i.e. isolation). Because pair housing is a form of enrichment (Council, 2011), comparing EC to pair-housed "control" rats would be akin to comparing a 20 mg/kg cocaine group to a 10 mg/kg "control". The problem is that "standard" laboratory housing for rats is two per cage in most research laboratories, and as such, these pair-housed rats are viewed by many scientists to constitute the "normal", or control condition. From this viewpoint, instead of seeing a continuum of enrichment ranging from isolation to pair-housed to full enrichment, environmental enrichment is seen as one manipulation (compared to pair-housed rats) and environmental isolation is seen as a different kind of manipulation.

At first sight, the case for isolation being a different kind of manipulation might seem to have merit. The case that many researchers make is that isolation itself is a stressor, and as such should not be used as a control. Indeed, many studies have shown that maternal separation (isolating pups from dam) and neonatal isolation (isolating pups from dam and other pups) are significant stressors, evoking ultrasonic vocalizations and inducing CORT release in pups (Hennessy and Weinberg, 1990; Kehne et al., 1991; Kuhn et al., 1990; Levine et al., 1991; McCormick et al., 1998; Viau et al., 1996). In addition, acute isolation of group-housed rats also induces CORT, which is taken as a clear indication of stress (Takatsu-Coleman et al., 2013). Many researchers thus make the leap of considering isolation rearing a condition of chronic stress, citing large CORT induction from acute stress as evidence. However, the inoculation stress hypothesis states that enriched animals are in a state of chronic very mild stress. As described above, *blunted* CORT induction is a sign of chronic stress, which is seen in EC rats and chronic mildly stressed animals alike. Table 1 further illustrates the many other signs of chronic stress that EC rats show that IC rats do not show. It is important to make the distinction that exhibiting a greater response to a stressor is not the same as being chronically stressed—quite the opposite. The isolated animals essentially have a lack of daily stimulation (i.e. stress) and therefore show a greater response to stress than EC rats, which are constantly stimulated (i.e. stressed) and therefore show a lesser response to subsequent stress.

Arguments frequently used in favor of pair-housed controls over isolated controls is that the IC group is not a "natural" condition, nor a "normal" condition, and has less "translational relevance" than pair-housed rats. Pair-housing rats in a laboratory is certainly not more translationally relevant than any other housing condition. The only place it is normal for two humans to be confined in the same small space is in prison. As for "normal", even wild rats are social animals and are found in groups rather than pairs that stick together constantly. Regardless, is a control group itself supposed to be "normal"? In the case of a pharmacological control group, would it be "normal" for a human to take an injection of saline rather than a drug? A control group should be the lack of a manipulation, but the pair-housed or social condition contains some of the variables of enrichment, namely social interaction and elements of exercise or play, and therefore is an intermediate level of enrichment capable of producing some behavioral effects of full enrichment, but not all (Bardo et al., 2001b; Green et al., 2010; Rosenzweig, 2003; Zakharova et al., 2012). For as described above in the inoculation stress hypothesis, environmental enrichment is a combination of exercise, novelty, and social contact with conspecifics, all of which are mild stressors resulting in inoculation against future stressors. Neither social interaction, novelty, nor any other single variable can account for all of the effects of enrichment (van Praag et al., 2000). However, the goal in using the environmental enrichment paradigm is not to tease apart the different aspects of enrichment or different gradations of enrichment but rather to determine differences between enrichment and the lack of enrichment to find the molecular determinants of the resilience to addiction and depression. Therefore, in our opinion in light of the inoculation stress hypothesis, the isolated condition is the correct control for enrichment and the inclusion of an intermediate group such as the social condition is unnecessary.

11. Conclusions

The inoculation stress hypothesis is an under-appreciated framework for understanding many of the complexities of stress and an organism's response to that stress. This hypothesis also has utility as a scaffold for which to build other novel hypotheses concerning susceptibility and resistance to psychiatric conditions such as addiction and major depression. Accordingly, application of the inoculation stress hypothesis to the environmental enrichment paradigm helps clarify the nature of nurture (i.e. role of environment) and its contribution to the resilience to addiction- and depression-related behavior. Future research designed from a better understanding of stress and environment will help to identify novel targets for the treatment of addiction and depression. Current research is already making headway (Fan et al., 2013a,b; Pavlovsky et al., 2013).

Acknowledgments

This research is funded by NIDA DA029091 and DA007287, and NINDS grant NS081121.

References

Abelson, K.S., Adem, B., Royo, F., Carlsson, H.E., Hau, J., 2005. High plasma corticosterone levels persist during frequent automatic blood sampling in rats. In Vivo (Athens, Greece) 19, 815–819.

Ahmed, S.H., Koob, G.F., 1997. Cocaine- but not food-seeking behavior is reinstated by stress after extinction. Psychopharmacology (Berl) 132, 289–295.

Alibhai, I.N., Green, T.A., Potashkin, J.A., Nestler, E.J., 2007. Regulation of fosB and DeltafosB mRNA expression: in vivo and in vitro studies. Brain Res. 1143, 22–33.

Allen, C., Kendall, J.W., 1967. Maturation of the circadian rhythm of plasma corticosterone in the rat. Endocrinology 80, 926–930.

Alvers, K.M., Marusich, J.A., Gipson, C.D., Beckmann, J.S., Bardo, M.T., 2012. Environmental enrichment during development decreases intravenous self-administration of methylphenidate at low unit doses in rats. Behav. Pharmacol. 23, 650–657.

Bardo, M., Klebaur, J., Valone, J., Deaton, C., 2001a. Environmental enrichment decreases intravenous self-administration of amphetamine in female and male rats. Psychopharmacology (Berl) 155, 278–284.

Bardo, M., Valone, J., Robinet, P., Shaw, W., Dwoskin, L., 1999. Environmental enrichment enhances the stimulant effect of intravenous amphetamine: search for a cellular mechanism in the nucleus accumbens. Psychobiology 27, 292–299.

Bardo, M.T., Bevins, R.A., 2000. Conditioned place preference: what does it add to our preclinical understanding of drug reward? Psychopharmacology (Berl) 153, 31–43.

Bardo, M.T., Bowling, S.L., Rowlett, J.K., Manderscheid, P., Buxton, S.T., Dwoskin, L.P., 1995. Environmental enrichment attenuates locomotor sensitization, but

not in vitro dopamine release, induced by amphetamine. Pharmacol. Biochem. Behav. 51, 397–405.
Bardo, M.T., Klebaur, J.E., Valone, J.M., Deaton, C., 2001b. Environmental enrichment decreases intravenous self-administration of amphetamine in female and male rats. Psychopharmacology (Berl) 155, 278–284.
Barnum, C.J., Blandino Jr., P., Deak, T., 2007. Adaptation in the corticosterone and hyperthermic responses to stress following repeated stressor exposure. J. Neuroendocrinol. 19, 632–642.
Belke, T.W., 2000. Studies of wheel-running reinforcement: parameters of Herrnstein's (1970) response-strength equation vary with schedule order. J. Exp. Anal. Behav. 73, 319–331.
Belz, E.E., Kennell, J.S., Czambel, R.K., Rubin, R.T., Rhodes, M.E., 2003. Environmental enrichment lowers stress-responsive hormones in singly housed male and female rats. Pharmacol. Biochem. Behav. 76, 481–486.
Bevins, R.A., Bardo, M.T., 1999. Conditioned increase in place preference by access to novel objects: antagonism by MK-801. Behav. Brain Res. 99, 53–60.
Bezard, E., Dovero, S., Belin, D., Duconger, S., Jackson-Lewis, V., Przedborski, S., Piazza, P.V., Gross, C.E., Jaber, M., 2003. Enriched environment confers resistance to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine and cocaine: involvement of dopamine transporter and trophic factors. J. Neurosci. 23, 10999–11007.
Bouet, V., Freret, T., Dutar, P., Billard, J.M., Boulouard, M., 2011. Continuous enriched environment improves learning and memory in adult NMRI mice through theta burst-related-LTP independent mechanisms but is not efficient in advanced aged animals. Mech. Ageing Dev. 132, 240–248.
Bowling, S., Bardo, M., 1994a. Locomotor and rewarding effects of amphetamine in enriched, social, and isolate reared rats. Pharmacol. Biochem. Behav. 48, 459–464.
Bowling, S.L., Bardo, M.T., 1994b. Locomotor and rewarding effects of amphetamine in enriched, social, and isolate reared rats. Pharmacol. Biochem. Behav. 48, 459–464.
Bowling, S.L., Rowlett, J.K., Bardo, M.T., 1993. The effect of environmental enrichment on amphetamine-stimulated locomotor activity, dopamine synthesis and dopamine release. Neuropharmacology 32, 885–893.
Brenes, J.C., Rodriguez, O., Fornaguera, J., 2008. Differential effect of environment enrichment and social isolation on depressive-like behavior, spontaneous activity and serotonin and norepinephrine concentration in prefrontal cortex and ventral striatum. Pharmacol. Biochem. Behav. 89, 85–93.
Brenes Saenz, J.C., Villagra, O.R., Fornaguera Trias, J., 2006. Factor analysis of forced swimming test, sucrose preference test and open field test on enriched, social and isolated reared rats. Behav. Brain Res. 169, 57–65.
Bruijnzeel, A.W., 2012. Tobacco addiction and the dysregulation of brain stress systems. Neurosci. Biobehav. Rev. 36, 1418–1441.
Buono, M.J., Yeager, J.E., Hodgdon, J.A., 1986. Plasma adrenocorticotropin and cortisol responses to brief high-intensity exercise in humans. J. Appl. Physiol. 61, 1337–1339.
Burke, A.R., Miczek, K.A., 2014. Stress in adolescence and drugs of abuse in rodent models: role of dopamine, CRF, and HPA axis. Psychopharmacology (Berl) 231, 1557–1580.
Butte, J.C., Kakihana, R., Noble, E.P., 1976. Circadian rhythm of corticosterone levels in rat brain. J. Endocrinol. 68, 235–239.
Buwalda, B., Scholte, J., de Boer, S.F., Coppens, C.M., Koolhaas, J.M., 2012. The acute glucocorticoid stress response does not differentiate between rewarding and aversive social stimuli in rats. Horm. Behav. 61, 218–226.
Carroll, M.E., Meisch, R.A., 1984. Increased drug-reinforced behavior due to food deprivation. Adv. Behav. Pharmacol. 4, 47–88.
Carter, R.N., Pinnock, S.B., Herbert, J., 2004. Does the amygdala modulate adaptation to repeated stress? Neuroscience 126, 9–19.
Chauvet, C., Goldberg, S.R., Jaber, M., Solinas, M., 2012. Effects of environmental enrichment on the incubation of cocaine craving. Neuropharmacology 63, 635–641.
Chen, X., Herbert, J., 1995. Regional changes in c-fos expression in the basal forebrain and brainstem during adaptation to repeated stress: correlations with cardiovascular, hypothermic and endocrine responses. Neuroscience 64, 675–685.
Council, N.R., 2011. Guide for the Care and Use of Laboratory Animals: Eighth Edition. The National Academies Press.
Covington 3rd, H.E., Miczek, K.A., 2005. Intense cocaine self-administration after episodic social defeat stress, but not after aggressive behavior: dissociation from corticosterone activation. Psychopharmacology (Berl) 183, 331–340.
Cruz, F.C., Marin, M.T., Leao, R.M., Planeta, C.S., 2012. Behavioral and neuroendocrine effects of the exposure to chronic restraint or variable stress in early adolescent rats. Int. J. Dev. Neurosci 30, 19–23.
De Boer, S.F., Koopmans, S.J., Slangen, J.L., Van der Gugten, J., 1990. Plasma catecholamine, corticosterone and glucose responses to repeated stress in rats: effect of interstressor interval length. Physiol. Behav. 47, 1117–1124.
Deehan Jr., G.A., Cain, M.E., Kiefer, S.W., 2007. Differential rearing conditions alter operant responding for ethanol in outbred rats. Alcohol. Clin. Exp. Res. 31, 1692–1698.
Deinzer, R., Kirschbaum, C., Gresele, C., Hellhammer, D.H., 1997. Adrenocortical responses to repeated parachute jumping and subsequent h-CRH challenge in inexperienced healthy subjects. Physiol. Behav. 61, 507–511.
Deroche, V., Piazza, P.V., Casolini, P., Maccari, S., Le Moal, M., Simon, H., 1992. Stress-induced sensitization to amphetamine and morphine psychomotor effects depend on stress-induced corticosterone secretion. Brain Res. 598, 343–348.
Deroche, V., Piazza, P.V., Deminiere, J.M., Le Moal, M., Simon, H., 1993. Rats orally self-administer corticosterone. Brain Res. 622, 315–320.
Diamond, M.C., Krech, D., Rosenzweig, M.R., 1964. The effects of an enriched environment on the histology of the rat cerebral cortex. J. Comp. Neurol. 123, 111–120.
Dienstbier, R.A., 1989. Arousal and physiological toughness: implications for mental and physical health. Psychol. Rev. 96, 84–100.
Dobrakovova, M., Kvetnansky, R., Oprsalova, Z., Macho, L., 1990. [The effect of chronic stress on the activity of the sympathetic-adrenomedullary system]. Bratisl. Lek. Listy 91, 587–592.
Erb, S., Shaham, Y., Stewart, J., 1996. Stress reinstates cocaine-seeking behavior after prolonged extinction and a drug-free period. Psychopharmacology (Berl) 128, 408–412.
Fan, X., Li, D., Lichti, C.F., Green, T.A., 2013a. Dynamic proteomics of nucleus accumbens in response to acute psychological stress in environmentally enriched and isolated rats. PLOS ONE 8, e73689.
Fan, X., Li, D., Zhang, Y., Green, T.A., 2013b. Differential phosphoproteome regulation of nucleus accumbens in environmentally enriched and isolated rats in response to acute stress. PLOS ONE 8, e79893.
Fediuc, S., Campbell, J.E., Riddell, M.C., 2006. Effect of voluntary wheel running on circadian corticosterone release and on HPA axis responsiveness to restraint stress in Sprague–Dawley rats. J. Appl. Physiol. 100, 1867–1875 (Bethesda, Md.: 1985).
Felitti, V.J., Anda, R.F., Nordenberg, D., Williamson, D.F., Spitz, A.M., Edwards, V., Koss, M.P., Marks, J.S., 1998. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. Am. J. Prev. Med. 14, 245–258.
Fox, C., Merali, Z., Harrison, C., 2006. Therapeutic and protective effect of environmental enrichment against psychogenic and neurogenic stress. Behav. Brain Res. 175, 1–8.
Friske, J.E., Gammie, S.C., 2005. Environmental enrichment alters plus maze, but not maternal defense performance in mice. Physiol. Behav. 85, 187–194.
Gadek-Michalska, A., Bugajski, J., 2003. Repeated handling, restraint, or chronic crowding impair the hypothalamic–pituitary–adrenocortical response to acute restraint stress. J. Physiol. Pharmacol. 54, 449–459.
Gärtner, K., Büttner, D., Döhler, K., Friedel, R., Lindena, J., Trautschold, I., 1980. Stress response of rats to handling and experimental procedures. Lab. Anim. 14, 267–274.
Gerra, G., Zaimovic, A., Mascetti, G.G., Gardini, S., Zambelli, U., Timpano, M., Raggi, M.A., Brambilla, F., 2001. Neuroendocrine responses to experimentally-induced psychological stress in healthy humans. Psychoneuroendocrinology 26, 91–107.
Gill, K.E., Beveridge, T.J., Smith, H.R., Porrino, L.J., 2013. The effects of rearing environment and chronic methylphenidate administration on behavior and dopamine receptors in adolescent rats. Brain Res. 1527, 67–78.
Goeders, N.E., 2002. Stress and cocaine addiction. J. Pharmacol. Exp. Ther. 301, 785–789.
Goeders, N.E., Guerin, G.F., 1994. Non-contingent electric footshock facilitates the acquisition of intravenous cocaine self-administration in rats. Psychopharmacology (Berl) 114, 63–70.
Goeders, N.E., Guerin, G.F., 1996a. Effects of surgical and pharmacological adrenalectomy on the initiation and maintenance of intravenous cocaine self-administration in rats. Brain Res. 722, 145–152.
Goeders, N.E., Guerin, G.F., 1996b. Role of corticosterone in intravenous cocaine self-administration in rats. Neuroendocrinology 64, 337–348.
Graf, E.N., Wheeler, R.A., Baker, D.A., Ebben, A.L., Hill, J.E., McReynolds, J.R., Robble, M.A., Vranjkovic, O., Wheeler, D.S., Mantsch, J.R., Gasser, P.J., 2013. Corticosterone acts in the nucleus accumbens to enhance dopamine signaling and potentiate reinstatement of cocaine seeking. J. Neurosci 33, 11800–11810.
Green, T.A., Alibhai, I.N., Hommel, J.D., DiLeone, R.J., Kumar, A., Theobald, D.E., Neve, R.L., Nestler, E.J., 2006. Induction of inducible cAMP early repressor expression in nucleus accumbens by stress or amphetamine increases behavioral responses to emotional stimuli. J. Neurosci 26, 8235–8242.
Green, T.A., Alibhai, I.N., Roybal, C.N., Winstanley, C.A., Theobald, D.E., Birnbaum, S.G., Graham, A.R., Unterberg, S., Graham, D.L., Vialou, V., Bass, C.E., Terwilliger, E.F., Bardo, M.T., Nestler, E.J., 2010. Environmental enrichment produces a behavioral phenotype mediated by low cyclic adenosine monophosphate response element binding (CREB) activity in the nucleus accumbens. Biol. Psychiatry 67, 28–35.
Green, T.A., Alibhai, I.N., Unterberg, S., Neve, R.L., Ghose, S., Tamminga, C.A., Nestler, E.J., 2008. Induction of activating transcription factors (ATFs) ATF2, ATF3, and ATF4 in the nucleus accumbens and their regulation of emotional behavior. J. Neurosci. 28, 2025–2032.
Green, T.A., Cain, M.E., Thompson, M., Bardo, M.T., 2003. Environmental enrichment decreases nicotine-induced hyperactivity in rats. Psychopharmacology (Berl) 170, 235–241.
Green, T.A., Gehrke, B.J., Bardo, M.T., 2002. Environmental enrichment decreases intravenous amphetamine self-administration in rats: dose–response functions for fixed- and progressive-ratio schedules. Psychopharmacology (Berl) 162, 373–378.
Greenwood, B.N., Foley, T.E., Le, T.V., Strong, P.V., Loughridge, A.B., Day, H.E., Fleshner, M., 2011. Long-term voluntary wheel running is rewarding and produces plasticity in the mesolimbic reward pathway. Behav. Brain Res. 217, 354–362.
Gregus, A., Wintink, A.J., Davis, A.C., Kalynchuk, L.E., 2005. Effect of repeated corticosterone injections and restraint stress on anxiety and depression-like behavior in male rats. Behav. Brain Res. 156, 105–114.
Grimm, J.W., Osincup, D., Wells, B., Manaois, M., Fyall, A., Buse, C., Harkness, J.H., 2008. Environmental enrichment attenuates cue-induced reinstatement of sucrose seeking in rats. Behav. Pharmacol. 19, 777–785.
Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W.A., Beekman, A.T., 2013. Recurrence of major depressive disorder and its predictors in the general population: results

from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). Psychol. Med. 43, 39–48.
Harlow, H.F., 1958. The nature of love. Am. Psychol. 13, 673.
Harlow, H.F., Suomi, S.J., 1971. Social recovery by isolation-reared monkeys. Proc. Natl. Acad. Sci. U. S. A. 68, 1534–1538.
Harris, R.B., Gu, H., Mitchell, T.D., Endale, L., Russo, M., Ryan, D.H., 2004. Increased glucocorticoid response to a novel stress in rats that have been restrained. Physiol. Behav. 81, 557–568.
Hatcher, J.P., Bell, D.J., Reed, T.J., Hagan, J.J., 1997. Chronic mild stress-induced reductions in saccharin intake depend upon feeding status. J. Psychopharmacol. (Oxford, England) 11, 331–338.
Heim, C., Owens, M.J., Plotsky, P.M., Nemeroff, C.B., 1997. The role of early adverse life events in the etiology of depression and posttraumatic stress disorder. Focus on corticotropin-releasing factor. Ann. N. Y. Acad. Sci. 821, 194–207.
Hennessy, M.B., Heybach, J.P., Vernikos, J., Levine, S., 1979. Plasma corticosterone concentrations sensitively reflect levels of stimulus intensity in the rat. Physiol. Behav. 22, 821–825.
Hennessy, M.M., Weinberg, J., 1990. Adrenocortical activity during conditions of brief social separation in preweaning rats. Behav. Neural Biol. 54, 42–55.
Hockly, E., Cordery, P.M., Woodman, B., Mahal, A., van Dellen, A., Blakemore, C., Lewis, C.M., Hannan, A.J., Bates, G.P., 2002. Environmental enrichment slows disease progression in R6/2 Huntington's disease mice. Ann. Neurol. 51, 235–242.
Jankowsky, J.L., Melnikova, T., Fadale, D.J., Xu, G.M., Slunt, H.H., Gonzales, V., Younkin, L.H., Younkin, S.G., Borchelt, D.R., Savonenko, A.V., 2005. Environmental enrichment mitigates cognitive deficits in a mouse model of Alzheimer's disease. J. Neurosci. 25, 5217–5224.
Jankowsky, J.L., Xu, G., Fromholt, D., Gonzales, V., Borchelt, D.R., 2003. Environmental enrichment exacerbates amyloid plaque formation in a transgenic mouse model of Alzheimer disease. J. Neuropathol. Exp. Neurol. 62, 1220–1227.
Jeong, Y.H., Kim, J.M., Yoo, J., Lee, S.H., Kim, H.S., Suh, Y.H., 2011. Environmental enrichment compensates for the effects of stress on disease progression in Tg2576 mice, an Alzheimer's disease model. J. Neurochem. 119, 1282–1293.
Jorge, E., Fernandez, J.A., Torres, R., Vergara, P., Martin, M.T., 2010. Functional changes induced by psychological stress are not enough to cause intestinal inflammation in Sprague–Dawley rats. Neurogastroenterol. Motil. 22, e241–e250.
Kehne, J.H., McCloskey, T.C., Baron, B.M., Chi, E.M., Harrison, B.L., Whitten, J.P., Palfreyman, M.G., 1991. NMDA receptor complex antagonists have potential anxiolytic effects as measured with separation-induced ultrasonic vocalizations. Eur. J. Pharmacol. 193, 283–292.
Kendler, K.S., Karkowski, L.M., Prescott, C.A., 1998. Stressful life events and major depression: risk period, long-term contextual threat, and diagnostic specificity. J. Nerv. Ment. Dis. 186, 661–669.
Kendler, K.S., Karkowski, L.M., Prescott, C.A., 1999. Causal relationship between stressful life events and the onset of major depression. Am. J. Psychiatry 156, 837.
Kessler, R.C., Chiu, W.T., Demler, O., Merikangas, K.R., Walters, E.E., 2005. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. Arch. Gen. Psychiatry 62, 617–627.
Khoshaba, D.M., Maddi, S.R., 1999. Early experiences in hardiness development. Consult. Psychol. J.: Pract. Res. 51, 106–116.
Konkle, A.T., Kentner, A.C., Baker, S.L., Stewart, A., Bielajew, C., 2010. Environmental-enrichment–related variations in behavioral, biochemical, and physiologic responses of Sprague–Dawley and long Evans rats. J. Am. Assoc. Lab. Anim. Sci.: JAALAS 49, 427.
Koolhaas, J., De Boer, S., De Rutter, A., Meerlo, P., Sgoifo, A., 1996. Social stress in rats and mice. Acta Physiol. Scand. Suppl. 640, 69–72.
Kosten, T.A., Miserendino, M.J., Kehoe, P., 2000. Enhanced acquisition of cocaine self-administration in adult rats with neonatal isolation stress experience. Brain Res. 875, 44–50.
Kozorovitskiy, Y., Gross, C.G., Kopil, C., Battaglia, L., McBreen, M., Stranahan, A.M., Gould, E., 2005. Experience induces structural and biochemical changes in the adult primate brain. Proc. Natl. Acad. Sci. U. S. A. 102, 17478–17482.
Kuhn, C.M., Pauk, J., Schanberg, S.M., 1990. Endocrine responses to mother-infant separation in developing rats. Dev. Psychobiol. 23, 395–410.
Larsson, F., Winblad, B., Mohammed, A.H., 2002. Psychological stress and environmental adaptation in enriched vs. impoverished housed rats. Pharmacol. Biochem. Behav. 73, 193–207.
Laviola, G., Hannan, A.J., Macri, S., Solinas, M., Jaber, M., 2008. Effects of enriched environment on animal models of neurodegenerative diseases and psychiatric disorders. Neurobiol. Dis. 31, 159–168.
Leal-Galicia, P., Saldívar-González, A., Morimoto, S., Arias, C., 2007. Exposure to environmental enrichment elicits differential hippocampal cell proliferation: role of individual responsiveness to anxiety. Dev. Neurobiol. 67, 395–405.
Lehmann, M.L., Herkenham, M., 2011. Environmental enrichment confers stress resiliency to social defeat through an infralimbic cortex-dependent neuroanatomical pathway. J. Neurosci. 31, 6159–6173.
Lepsch, L., Gonzalo, L., Magro, F., Delucia, R., Scavone, C., Planeta, C., 2005. Exposure to chronic stress increases the locomotor response to cocaine and the basal levels of corticosterone in adolescent rats. Addict. Biol. 10, 251–256.
Levi, O., Jongen-Relo, A.L., Feldon, J., Roses, A.D., Michaelson, D.M., 2003. ApoE4 impairs hippocampal plasticity isoform-specifically and blocks the environmental stimulation of synaptogenesis and memory. Neurobiol. Dis. 13, 273–282.
Levin, F.R., Bisaga, A., Raby, W., Aharonovich, E., Rubin, E., Mariani, J., Brooks, D.J., Garawi, F., Nunes, E.V., 2008. Effects of major depressive disorder and attention-deficit/hyperactivity disorder on the outcome of treatment for cocaine dependence. J. Subst. Abuse Treat. 34, 80–89.
Levine, S., Huchton, D.M., Wiener, S.G., Rosenfeld, P., 1991. Time course of the effect of maternal deprivation on the hypothalamic–pituitary–adrenal axis in the infant rat. Dev. Psychobiol. 24, 547–558.
Lichti, C.F., Fan, F., English, R.D., Zhang, Y., Li, D., Kong, F., Sinha, M., Andersen, C.R., Spratt, H., Luxon, B.A., Green, T.A., 2014. Environmental enrichment alters protein expression as well as the proteomic response to cocaine in rat nucleus accumbens. Front. Behav. Neurosci. 8, 246, http://dx.doi.org/10.3389/fnbeh.2014.00246.
Lobo, M.K., Zaman, S., Damez-Werno, D.M., Koo, J.W., Bagot, R.C., DiNieri, J.A., Nugent, A., Finkel, E., Chaudhury, D., Chandra, R., Riberio, E., Rabkin, J., Mouzon, E., Cachope, R., Cheer, J.F., Han, M.H., Dietz, D.M., Self, D.W., Hurd, Y.L., Vialou, V., Nestler, E.J., 2013. DeltaFosB induction in striatal medium spiny neuron subtypes in response to chronic pharmacological, emotional, and optogenetic stimuli. J. Neurosci. 33, 18381–18395.
Louilot, A., Le Moal, M., Simon, H., 1986. Differential reactivity of dopaminergic neurons in the nucleus accumbens in response to different behavioral situations. An in vivo voltammetric study in free moving rats. Brain Res. 397, 395–400.
Lyons, D.M., Parker, K.J., Katz, M., Schatzberg, A.F., 2009. Developmental cascades linking stress inoculation, arousal regulation, and resilience. Front. Behav. Neurosci. 3, 32.
Lyons, D.M., Parker, K.J., Schatzberg, A.F., 2010. Animal models of early life stress: implications for understanding resilience. Dev. Psychobiol. 52, 402–410.
MacDougall, M.J., Howland, J.G., 2013. Acute stress, but not corticosterone, disrupts short- and long-term synaptic plasticity in rat dorsal subiculum via glucocorticoid receptor activation. Cereb. Cortex (New York, N.Y.: 1991) 23, 2611–2619.
Magarinos, A.M., McEwen, B.S., 1995. Stress-induced atrophy of apical dendrites of hippocampal CA3c neurons: comparison of stressors. Neuroscience 69, 83–88.
Marti, O., Gavalda, A., Jolin, T., Armario, A., 1993. Effect of regularity of exposure to chronic immobilization stress on the circadian pattern of pituitary adrenal hormones, growth hormone, and thyroid stimulating hormone in the adult male rat. Psychoneuroendocrinology 18, 67–77.
Matthews, K., Wilkinson, L.S., Robbins, T.W., 1996. Repeated maternal separation of preweanling rats attenuates behavioral responses to primary and conditioned incentives in adulthood. Physiol. Behav. 59, 99–107.
McCormick, C.M., Kehoe, P., Kovacs, S., 1998. Corticosterone release in response to repeated, short episodes of neonatal isolation: evidence of sensitization. Int. J. Dev. Neurosci. 16, 175–185.
McEwen, B.S., 2000. Allostasis and allostatic load: implications for neuropsychopharmacology. Neuropsychopharmacology 22, 108–124.
McEwen, B.S., Wingfield, J.C., 2010. What is in a name? Integrating homeostasis, allostasis and stress. Horm. Behav. 57, 105–111.
Meichenbaum, D., 2007. Stress inoculation training: a preventative and treatment approach. In: Lehrer, P.M., Woolfolk, R.L., Sime, W.S. (Eds.), Principles and Practice of Stress Management. , third ed. Guilford Press.
Miczek, K.A., Mutschler, N.H., 1996. Activational effects of social stress on IV cocaine self-administration in rats. Psychopharmacology (Berl) 128, 256–264.
Mizoguchi, K., Yuzurihara, M., Ishige, A., Sasaki, H., Chui, D.H., Tabira, T., 2001. Chronic stress differentially regulates glucocorticoid negative feedback response in rats. Psychoneuroendocrinology 26, 443–459.
Mlynarik, M., Johansson, B.B., Jezova, D., 2004. Enriched environment influences adrenocortical response to immune challenge and glutamate receptor gene expression in rat hippocampus. Ann. N.Y. Acad. Sci. 1018, 273–280.
Moffett, M.C., Harley, J., Francis, D., Sanghani, S.P., Davis, W.I., Kuhar, M.J., 2006. Maternal separation and handling affects cocaine self-administration in both the treated pups as adults and the dams. J. Pharmacol. Exp. Ther. 317, 1210–1218.
Moldow, R.L., Fischman, A.J., 1987. Cocaine induced secretion of ACTH, beta-endorphin, and corticosterone. Peptides 8, 819–822.
Moncek, F., Duncko, R., Johansson, B., Jezova, D., 2004. Effect of environmental enrichment on stress related systems in rats. J. Neuroendocrinol. 16, 423–431.
Morris, M.C., Ciesla, J.A., Garber, J., 2010. A prospective study of stress autonomy versus stress sensitization in adolescents at varied risk for depression. J. Abnorm. Psychol. 119, 341–354.
Morris, M.C., Rao, U., Garber, J., 2012. Cortisol responses to psychosocial stress predict depression trajectories: social-evaluative threat and prior depressive episodes as moderators. J. Affect. Disord. 143, 223–230.
Mortimer, J.T., Staff, J., 2004. Early work as a source of developmental discontinuity during the transition to adulthood. Dev. Psychopathol. 16, 1047–1070.
Müller, M.B., Landgraf, R., Preil, J., Sillaber, I., Kresse, A.E., Keck, M.E., Zimmermann, S., Holsboer, F., Wurst, W., 2000. Selective activation of the hypothalamic vasopressinergic system in mice deficient for the corticotropin-releasing hormone receptor 1 is dependent on glucocorticoids 1. Endocrinology 141, 4262–4269.
Murison, R., Hansen, A.L., 2001. Reliability of the chronic mild stress paradigm: implications for research and animal welfare. Integr. Physiol. Behav. Sci. 36, 266–274.
Nader, J., Claudia, C., El Rawas, R., Favot, L., Jaber, M., Thiriet, N., Solinas, M., 2012. Loss of environmental enrichment increases vulnerability to cocaine addiction. Neuropsychopharmacology 37, 1579–1587.
Natelson, B.H., Ottenweller, J.E., Cook, J.A., Pitman, D., McCarty, R., Tapp, W.N., 1988. Effect of stressor intensity on habituation of the adrenocortical stress response. Physiol. Behav. 43, 41–46.
Ostrander, M.M., Ulrich-Lai, Y.M., Choi, D.C., Richtand, N.M., Herman, J.P., 2006. Hypoactivity of the hypothalamo-pituitary-adrenocortical axis during recovery from chronic variable stress. Endocrinology 147, 2008–2017.

Parker, K.J., Buckmaster, C.L., Schatzberg, A.F., Lyons, D.M., 2004. Prospective investigation of stress inoculation in young monkeys. Arch. Gen. Psychiatry 61, 933–941.
Parker, K.J., Maestripieri, D., 2011. Identifying key features of early stressful experiences that produce stress vulnerability and resilience in primates. Neurosci. Biobehav. Rev. 35, 1466–1483.
Pavlovsky, A.A., Boehning, D., Li, D., Zhang, Y., Fan, X., Green, T.A., 2013. Psychological stress, cocaine and natural reward each induce endoplasmic reticulum stress genes in rat brain. Neuroscience 246, 160–169.
Pena, Y., Prunell, M., Rotllant, D., Armario, A., Escorihuela, R.M., 2009. Enduring effects of environmental enrichment from weaning to adulthood on pituitary–adrenal function, pre-pulse inhibition and learning in male and female rats. Psychoneuroendocrinology 34, 1390–1404.
Perrotti, L.I., Hadeishi, Y., Ulery, P.G., Barrot, M., Monteggia, L., Duman, R.S., Nestler, E.J., 2004. Induction of deltaFosB in reward-related brain structures after chronic stress. J. Neurosci. 24, 10594–10602.
Pettinati, H.M., O'Brien, C.P., Dundon, W.D., 2013. Current status of co-occurring mood and substance use disorders: a new therapeutic target. Am. J. Psychiatry 170, 23–30.
Piazza, P.V., Deminiere, J.M., le Moal, M., Simon, H., 1990. Stress- and pharmacologically-induced behavioral sensitization increases vulnerability to acquisition of amphetamine self-administration. Brain Res. 514, 22–26.
Piazza, P.V., Le Moal, M., 1998. The role of stress in drug self-administration. Trends Pharmacol. Sci. 19, 67–74.
Piazza, P.V., Maccari, S., Deminière, J.M., Le Moal, M., Mormède, P., Simon, H., 1991. Corticosterone levels determine individual vulnerability to amphetamine self-administration. Proc. Natl. Acad. Sci. U. S. A. 88, 2088–2092.
Pohorecky, L.A., 1991. Stress and alcohol interaction: an update of human research. Alcohol. Clin. Exp. Res. 15, 438–459.
Puhl, M.D., Blum, J.S., Acosta-Torres, S., Grigson, P.S., 2012. Environmental enrichment protects against the acquisition of cocaine self-administration in adult male rats, but does not eliminate avoidance of a drug-associated saccharin cue. Behav. Pharmacol. 23, 43–53.
Rabasa, C., Delgado-Morales, R., Munoz-Abellan, C., Nadal, R., Armario, A., 2011a. Adaptation of the hypothalamic–pituitary–adrenal axis and glucose to repeated immobilization or restraint stress is not influenced by associative signals. Behav. Brain Res. 217, 232–239.
Rabasa, C., Munoz-Abellan, C., Daviu, N., Nadal, R., Armario, A., 2011b. Repeated exposure to immobilization or two different footshock intensities reveals differential adaptation of the hypothalamic–pituitary–adrenal axis. Physiol. Behav. 103, 125–133.
Ranaldi, R., Kest, K., Zellner, M., Hachimine-Semprebom, P., 2011. Environmental enrichment, administered after establishment of cocaine self-administration, reduces lever pressing in extinction and during a cocaine context renewal test. Behav. Pharmacol. 22, 347–353.
Raz, S., 2013. Ameliorative effects of brief daily periods of social interaction on isolation-induced behavioral and hormonal alterations. Physiol. Behav. 116–117, 13–22.
Rebec, G.V., Grabner, C.P., Johnson, M., Pierce, R.C., Bardo, M.T., 1997. Transient increases in catecholaminergic activity in medial prefrontal cortex and nucleus accumbens shell during novelty. Neuroscience 76, 707–714.
Renner, M.J., Rosenzweig, M.R., 1987. Enriched and Impoverished Environments: Effects on Brain and Behavior. Springer, New York.
Retana-Marquez, S., Bonilla-Jaime, H., Velazquez-Moctezuma, J., 1998. Lack of effect of corticosterone administration on male sexual behavior of rats. Physiol. Behav. 63, 367–370.
Rojas Vega, S., Strüder, H.K., Vera Wahrmann, B., Schmidt, A., Bloch, W., Hollmann, W., 2006. Acute BDNF and cortisol response to low intensity exercise and following ramp incremental exercise to exhaustion in humans. Brain Res. 1121, 59–65.
Rosenzweig, M.R., 2003. Effects of differential experience on the brain and behavior. Dev. Neuropsychol. 24, 523–540.
Rosenzweig, M.R., Bennett, E.L., 1996. Psychobiology of plasticity: effects of training and experience on brain and behavior. Behav. Brain Res. 78, 57–65.
Roy, V., Belzung, C., Delarue, C., Chapillon, P., 2001. Environmental enrichment in BALB/c mice: effects in classical tests of anxiety and exposure to a predatory odor. Physiol. Behav. 74, 313–320.
Rutter, M., 2006. Implications of resilience concepts for scientific understanding. Ann. N.Y. Acad. Sci. 1094, 1–12.
Seery, M.D., Leo, R.J., Lupien, S.P., Kondrak, C.L., Almonte, J.L., 2013. An upside to adversity?: moderate cumulative lifetime adversity is associated with resilient responses in the face of controlled stressors. Psychol. Sci. 24, 1181–1189.
Selye, H., 1976. Stress without distress. In: Serban, G. (Ed.), Psychopathology of Human Adaptation. Springer, New York, pp. 137–146.
Shaham, Y., Stewart, J., 1994. Exposure to mild stress enhances the reinforcing efficacy of intravenous heroin self-administration in rats. Psychopharmacology (Berl) 114, 523–527.
Sharp, J.L., Zammit, T.G., Azar, T.A., Lawson, D.M., 2002. Stress-like responses to common procedures in male rats housed alone or with other rats. Contemp. Top. Lab. Anim. Sci./Am. Assoc. Lab. Anim. Sci. 41, 8–14.
Shoji, H., Mizoguchi, K., 2010. Acute and repeated stress differentially regulates behavioral, endocrine, neural parameters relevant to emotional and stress response in young and aged rats. Behav. Brain Res. 211, 169–177.
Sih, A., 2011. Effects of early stress on behavioral syndromes: an integrated adaptive perspective. Neurosci. Biobehav. Rev. 35, 1452–1465.
Sinha, R., 2001. How does stress increase risk of drug abuse and relapse? Psychopharmacology (Berl) 158, 343–359.
Sinha, R., Catapano, D., O'Malley, S., 1999. Stress-induced craving and stress response in cocaine dependent individuals. Psychopharmacology (Berl) 142, 343–351.
Sinha, R., Garcia, M., Paliwal, P., Kreek, M.J., Rounsaville, B.J., 2006. Stress-induced cocaine craving and hypothalamic–pituitary–adrenal responses are predictive of cocaine relapse outcomes. Arch. Gen. Psychiatry 63, 324–331.
Skwara, A.J., Karwoski, T.E., Czambel, R.K., Rubin, R.T., Rhodes, M.E., 2012. Influence of environmental enrichment on hypothalamic–pituitary–adrenal (HPA) responses to single-dose nicotine, continuous nicotine by osmotic mini-pumps, and nicotine withdrawal by mecamylamine in male and female rats. Behav. Brain Res. 234, 1–10.
Smith, J.K., Neill, J.C., Costall, B., 1997. Post-weaning housing conditions influence the behavioural effects of cocaine and D-amphetamine. Psychopharmacology (Berl) 131, 23–33.
Smith, M.A., Iordanou, J.C., Cohen, M.B., Cole, K.T., Gergans, S.R., Lyle, M.A., Schmidt, K.T., 2009. Effects of environmental enrichment on sensitivity to cocaine in female rats: importance of control rates of behavior. Behav. Pharmacol. 20, 312–321.
Solinas, M., Chauvet, C., Thiriet, N., El Rawas, R., Jaber, M., 2008. Reversal of cocaine addiction by environmental enrichment. Proc. Natl. Acad. Sci. U. S. A. 105, 17145–17150.
Solinas, M., Thiriet, N., Chauvet, C., Jaber, M., 2010. Prevention and treatment of drug addiction by environmental enrichment. Prog. Neurobiol. 92, 572–592.
Solinas, M., Thiriet, N., El Rawas, R., Lardeux, V., Jaber, M., 2009. Environmental enrichment during early stages of life reduces the behavioral, neurochemical, and molecular effects of cocaine. Neuropsychopharmacology 34, 1102–1111.
Spires, T.L., Grote, H.E., Varshney, N.K., Cordery, P.M., van Dellen, A., Blakemore, C., Hannan, A.J., 2004. Environmental enrichment rescues protein deficits in a mouse model of Huntington's disease, indicating a possible disease mechanism. J. Neurosci. 24, 2270–2276.
Stairs, D.J., Bardo, M.T., 2009. Neurobehavioral effects of environmental enrichment and drug abuse vulnerability. Pharmacol. Biochem. Behav. 92, 377–382.
Stairs, D.J., Klein, E.D., Bardo, M.T., 2006. Effects of environmental enrichment on extinction and reinstatement of amphetamine self-administration and sucrose-maintained responding. Behav. Pharmacol. 17, 597–604.
Stairs, D.J., Prendergast, M.A., Bardo, M.T., 2011. Environmental-induced differences in corticosterone and glucocorticoid receptor blockade of amphetamine self-administration in rats. Psychopharmacology (Berl) 218, 293–301.
Swanson, H.H., van de Poll, N.E., 1983. Effects of an isolated or enriched environment after handling on sexual maturation and behaviour in male and female rats. J. Reprod. Fertil. 69, 165–171.
Takatsu-Coleman, A.L., Patti, C.L., Zanin, K.A., Zager, A., Carvalho, R.C., Borcoi, A.R., Ceccon, L.M., Berro, L.F., Tufik, S., Andersen, M.L., Frussa-Filho, R., 2013. Short-term social isolation induces depressive-like behaviour and reinstates the retrieval of an aversive task: mood-congruent memory in male mice? J. Psychiatry Neurosci.: JPN 38, 259–268.
Tennant, C., 2001. Work-related stress and depressive disorders. J. Psychosom. Res. 51, 697–704.
Thiel, K.J., Pentkowski, N.S., Peartree, N.A., Painter, M.R., Neisewander, J.L., 2010. Environmental living conditions introduced during forced abstinence alter cocaine-seeking behavior and Fos protein expression. Neuroscience 171, 1187–1196.
Thiel, K.J., Sanabria, F., Pentkowski, N.S., Neisewander, J.L., 2009. Anti-craving effects of environmental enrichment. Int. J. Neuropsychopharmacol. 12, 1151–1156.
Torres, G., Rivier, C., 1992. Cocaine-induced stimulation of the rat hypothalamic–pituitary–adrenal axis is progressively attenuated following hourly-interval regimens of the drug. Life Sci. 51, 1041–1048.
Urakawa, S., Mitsushima, D., Shimozuru, M., Sakuma, Y., Kondo, Y., 2014. An enriched rearing environment calms adult male rat sexual activity: implication for distinct serotonergic and hormonal responses to females. PLOS ONE 9, e87911.
Urakawa, S., Takamoto, K., Hori, E., Sakai, N., Ono, T., Nishijo, H., 2013. Rearing in enriched environment increases parvalbumin-positive small neurons in the amygdala and decreases anxiety-like behavior of male rats. BMC Neurosci. 14, 13.
van Dellen, A., Blakemore, C., Deacon, R., York, D., Hannan, A.J., 2000. Delaying the onset of Huntington's in mice. Nature 404, 721–722.
van Praag, H., Kempermann, G., Gage, F.H., 2000. Neural consequences of environmental enrichment. Nat. Rev. Neurosci. 1, 191–198.
Vialou, V., Maze, I., Renthal, W., LaPlant, Q.C., Watts, E.L., Mouzon, E., Ghose, S., Tamminga, C.A., Nestler, E.J., 2010. Serum response factor promotes resilience to chronic social stress through the induction of DeltaFosB. J. Neurosci. 30, 14585–14592.
Viau, V., Sharma, S., Meaney, M.J., 1996. Changes in plasma adrenocorticotropin, corticosterone, corticosteroid-binding globulin, and hippocampal glucocorticoid receptor occupancy/translocation in rat pups in response to stress. J. Neuroendocrinol. 8, 1–8.
Wallace, D.L., Han, M.H., Graham, D.L., Green, T.A., Vialou, V., Iniguez, S.D., Cao, J.L., Kirk, A., Chakravarty, S., Kumar, A., Krishnan, V., Neve, R.L., Cooper, D.C., Bolanos, C.A., Barrot, M., McClung, C.A., Nestler, E.J., 2009. CREB regulation of nucleus accumbens excitability mediates social isolation-induced behavioral deficits. Nat. Neurosci. 12, 200–209.
Willner, P., 1997. Validity, reliability and utility of the chronic mild stress model of depression: a 10-year review and evaluation. Psychopharmacology (Berl) 134, 319–329.

Willner, P., Muscat, R., Papp, M., 1992. Chronic mild stress-induced anhedonia: a realistic animal model of depression. Neurosci. Biobehav. Rev. 16, 525–534.
Willner, P., Towell, A., Sampson, D., Sophokleous, S., Muscat, R., 1987. Reduction of sucrose preference by chronic unpredictable mild stress, and its restoration by a tricyclic antidepressant. Psychopharmacology (Berl) 93, 358–364.
Wolf, S.A., Kronenberg, G., Lehmann, K., Blankenship, A., Overall, R., Staufenbiel, M., Kempermann, G., 2006. Cognitive and physical activity differently modulate disease progression in the amyloid precursor protein (APP)-23 model of Alzheimer's disease. Biol. Psychiatry 60, 1314–1323.
Worley, M.J., Trim, R.S., Roesch, S.C., Mrnak-Meyer, J., Tate, S.R., Brown, S.A., 2012. Comorbid depression and substance use disorder: longitudinal associations between symptoms in a controlled trial. J. Subst. Abuse Treat. 43, 291–302.
Wright, R.L., Conrad, C.D., 2008. Enriched environment prevents chronic stress-induced spatial learning and memory deficits. Behav. Brain Res. 187, 41–47.
Zakharova, E., Starosciak, A., Wade, D., Izenwasser, S., 2012. Sex differences in the effects of social and physical environment on novelty-induced exploratory behavior and cocaine-stimulated locomotor activity in adolescent rats. Behav. Brain Res. 230, 92–99.
Zhang, Y., Crofton, E.J., Li, D., Lobo, M.K., Fan, X., Nestler, E.J., Green, T.A., 2014. Overexpression of DeltaFosB in nucleus accumbens mimics the protective addiction phenotype, but not the protective depression phenotype of environmental enrichment. Front. Behav. Neurosci., 8.

OPEN

Citation: Transl Psychiatry (2017) 7, e1105; doi:10.1038/tp.2017.40

www.nature.com/tp

ORIGINAL ARTICLE

Instability of default mode network connectivity in major depression: a two-sample confirmation study

T Wise[1,2,6], L Marwood[1,2,6], AM Perkins[1,2], A Herane-Vives[1,3], R Joules[4], DJ Lythgoe[4], W-M Luh[5], SCR Williams[2,4], AH Young[1,2], AJ Cleare[1,2,7] and D Arnone[1,2,7]

Major depression is associated with altered static functional connectivity in various brain networks, particularly the default mode network (DMN). Dynamic functional connectivity is a novel tool with little application in affective disorders to date, and holds the potential to unravel fluctuations in connectivity strength over time in major depression. We assessed stability of connectivity in major depression between the medial prefrontal cortex (mPFC) and posterior cingulate cortex (PCC), key nodes in the DMN that are implicated in ruminative cognitions. Functional connectivity stability between the mPFC and PCC over the course of a resting-state functional magnetic resonance imaging (fMRI) scan was compared between medication-free patients with major depression and healthy controls matched for age, sex and handedness. We tested replicability of the results in an independent sample using multi-echo resting-state fMRI. The primary sample included 20 patients and 19 controls, while the validation sample included 19 patients and 19 controls. Greater connectivity variability was detected in major depression between mPFC and PCC. This was demonstrated in both samples indicating that the results were reliable and were not influenced by the fMRI acquisition approach used. Our results demonstrate that alterations within the DMN in major depression go beyond changes in connectivity strength and extend to reduced connectivity stability within key DMN regions. Findings were robustly replicated across two independent samples. Further research is necessary to better understand the nature of these fluctuations in connectivity and their relationship to the aetiology of major depression.

Translational Psychiatry (2017) **7,** e1105; doi:10.1038/tp.2017.40; published online 25 April 2017

INTRODUCTION

Depression is a common illness with substantial negative consequences for sufferers and society.[1,2] A better insight into neurobiological changes contributing to symptom generation is a research priority to improve diagnosis and treatment.[3]

Neuroimaging has enhanced our understanding of the neurobiological mechanisms underlying depressive symptoms[3–6] by identifying potential alterations in the structural and functional brain networks.[4,7] The default mode network (DMN) is one system that has attracted great research interest in major depression. One reason for this relates to its importance in the generation of self-referential thoughts, negative rumination and depressive symptoms.[8] A recent meta-analysis of static connectivity studies in major depression demonstrated hyperconnectivity within the DMN and between the DMN and fronto-parietal systems.[4] Within the DMN, the subsystem connecting the medial prefrontal cortex (mPFC) with the posterior cingulate cortex (PCC) is considered pivotal in generating affective, self-directed judgements and thoughts.[8,9] Although heightened static functional connectivity has been shown in major depression within this subsystem in association with a ruminative cognitive style,[4,10] there is uncertainty about its functional temporal stability. Connectivity variability is a plausible mechanism in major depression to explain brain responses associated with cognitive demands and processing of emotions.[11]

There is, however, little dynamic functional connectivity research published to date[11,12] with only one study in major depression.[13] This study showed increased variability in connectivity within the DMN between the mPFC and the insula, which correlated with a ruminative thinking style and coexisted with decreased variability between the mPFC and the parahippocampal gyrus.[13] No effect was found between the mPFC and PCC in this study, despite these being crucial in the generation of self-directed negative affective cognitions, potentially due to the less powerful whole-brain analysis method used.

Hence, we specifically evaluated temporal connectivity variability between the mPFC and the PCC, a subsystem within the DMN, given its relevance to ruminative cognitions associated with depression.[8,9] Kaiser *et al.*[13] observed increased connectivity variability between the mPFC and insula, part of the 'salience network' that is known to influence synergistically this DMN subsystem when processing internally generated salient information,[14] correlating with levels of rumination. Based on this finding that connections related to rumination show increased variability, we hypothesized increased connectivity variability in major depression between the mPFC and the PCC, correlating

[1]Centre for Affective Disorders, Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [2]South London and Maudsley NHS Foundation Trust, London, UK; [3]Departamento de Clínicas, Facultad de Medicina, Universidad Católica del Norte, Coquimbo, Chile; [4]Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK and [5]Cornell MRI Facility, Cornell University, Ithaca, New York, NY, USA. Correspondence: T Wise, Centre for Affective Disorders, Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, 103 Denmark Hill, London SE5 8AF, UK.
E-mail: toby.wise@kcl.ac.uk
[6]These authors contributed equally to this work.
[7]These authors contributed equally to this work.
Received 11 June 2016; revised 14 December 2016; accepted 9 January 2017

with ruminative cognitions. We tested this hypothesis first in a primary sample of medication-free participants with major depression selected to be free from psychiatric comorbidity. We then validated the robustness of this result by replicating the findings in an independent clinical sample. As a further test of the stability of the results in the presence of significant clinical heterogeneity, the validation sample included patients with major depression and comorbid anxiety disorders. Furthermore, the robustness of the neuroimaging results to non-neural artefacts was ensured by utilizing multi-echo functional magnetic resonance imaging (fMRI), a recent development that is superior to traditional de-noising methods.[15]

MATERIALS AND METHODS

Participants

The right-handed participants aged 18–65 were recruited from the local community using online advertisements[16] and waiting lists of local psychological therapy services. Given the novelty of the method, we were unable to determine an expected effect size *a priori*. However, the sample size was chosen to be consistent with other studies in the area and we used a validation sample to ensure our results were robust. All patients met Diagnostic and Statistical Manual for Mental Disorders IV criteria for unipolar major depression (current or recurrent episode), as determined by clinical interview based on the Mini International Neuropsychiatric Interview.[17] In the primary sample, comorbid conditions were excluded. For the validation sample, comorbid anxiety disorders were allowed alongside major depression. Depression severity was assessed with the Montgomery–Åsberg Depression Rating Scale[18] and a score $\geqslant 18$ was required for inclusion. Raters for both samples were trained on an independent sample of patients and demonstrated high inter-rater reliability (Intraclass correlation coefficient = 0.96, $P = 0.004$). The severity of anxiety symptoms was assessed using the Hamilton Depression Rating Scale 17 anxiety/somatization factor items (anxiety subscale)[19,20] and handedness with the Edinburgh Handedness Inventory.[21] Trait rumination was assessed using the Rumination Response Scale (RRS),[22] a 22-item self-report measure. Patients were not receiving any form of treatment, psychological or pharmacological, at the time of scanning and were medication-free for $\geqslant 2$ weeks ($\geqslant 4$ weeks for fluoxetine) before MRI scanning. No subjects had been receiving treatment with medication requiring a longer washout period. Healthy controls were assessed to exclude personal and familial (first-degree relatives) psychiatric history. Exclusion criteria for all the subjects included history of head injury, illicit substance use in the preceding two months, unstable medical illness, any treatment with potential psychotropic properties or interference with participants' safety or data interpretation, pregnancy or other contraindications for scanning.

Ethics approval

The research was approved by the local ethics committee. The subjects provided written informed consent and were compensated financially for participating.

fMRI acquisition

The data for each sample were acquired on two identical GE MR750 3-Tesla scanners with 12-channel radiofrequency head coils. The participants fixated on a cross with their eyes open for the scan duration. For the primary sample, a 6-min resting-state scan using a T2*-weighted echo-planar imaging sequence was acquired (repetition time = 2000, echo time = 30 ms, field of view = 22.1 cm, flip angle = 75°, 39 slices, resolution = 3.3 mm^3). The cardiac signals and respiratory information were also recorded. For the validation sample, the data were acquired using an 8-min multi-echo sequence (repetition time = 2300 ms, echo time = 12.7/31/48 ms, field of view = 24 cm, flip angle = 90°, 33 slices, resolution = $3.75 \times 3.75 \times 4.2$ mm). An identical high-resolution T1-weighted structural image was acquired for both the samples.

fMRI preprocessing

The data were pre-processed with custom Nipype (http://nipy.org/nipype/) scripts, using tools from SPM12 (http://www.fil.ion.ucl.ac.uk/spm), FSL 5.0.9 (http://fsl.fmrib.ox.ac.uk/), AFNI (https://afni.nimh.nih.gov/afni/), along with custom code (available upon request). The first four volumes of the functional series were discarded to allow for equilibration effects. Slice timing correction was performed and the images were realigned and co-registered to the structural image using the normalized mutual information method in SPM12. For the primary sample, physiological signals (cardiac and respiratory) were regressed from the data using AFNI's RETROICOR[23] tool. For the validation sample, multi-echo data were pre-processed using the multi-echo independent component analysis tool in AFNI[15] to isolate components in the signal representing true blood oxygen level dependent (BOLD) signal. This was used in place of RETROICOR as it has been shown to be a more effective method of de-noising.[15] The remaining processing steps were identical for both samples for consistency. Six motion parameters (three translation, three rotation, determined from the middle echo image for the validation sample) plus time series extracted from the white matter and cerebrospinal fluid regions were regressed out of the data. Data were then temporally filtered from 0.008 to 0.09 Hz before being demeaned, de-trended and smoothed with a 6 mm full width at half maximum kernel. Thus, preprocessing for both samples was identical except for the method of de-noising used.

Motion scrubbing

As even minimal head motion can affect correlations calculated from resting-state data when not controlled for,[24] time points exhibiting excessive motion were scrubbed from the BOLD time series.[24] Motion at each time point was assessed using root mean square (RMS) intensity difference between volumes (REFRMS) and DVARS[24] as calculated using the FSL motion outliers tool with default thresholds. As directly removing time points would affect the length of the sliding window, and hence dynamic connectivity estimates, we instead interpolated time points showing substantial motion using third-order b-spline interpolation. To compare motion estimates between samples, we used both total distance travelled and framewise displacement.[24] All the analyses were performed on the scrubbed, pre-processed data.

Region of interest definition

We performed group canonical independent component analysis,[25] implemented in Nilearn and using 20 clusters, to identify the DMN components. The clusters centred on the posterior cingulate and mPFC regions in the DMN component were used to create regions of interest (ROIs) for the connectivity variability analysis. A 10 mm diameter sphere was created based on the peak of each cluster in the independent component analysis, and mean time series were extracted from each of these ROIs. This procedure was performed independently for the two samples to identify sample-specific ROIs. These ROIs were used in place of the entire cluster to provide a more consistent signal and avoid contamination from surrounding areas. To ensure that results were specific to these regions rather than being a global pattern, or caused by non-neural factors, we performed a negative control analysis between these regions and a 10 mm spherical ROI in the medial primary motor cortex (Montreal Neurological Institute coordinates = −1, −8, 63), a region not previously linked to depression.

Sliding window correlation analysis

The sliding window analysis was performed using custom Python (https://www.python.org/) scripts. The data were split into 40 s Gaussian moving windows, staggered by one repetition time, created using a Gaussian kernel with a standard deviation of 8 s (see Supplementary Methods for a detailed discussion of the sliding window methodology). This time period has been shown to be appropriate for characterizing dynamic functional connectivity[26] and provides a fine-grained picture of temporal changes in connectivity. For each window, correlations were computed between variance-normalized time series from the two regions using Pearson correlations, the results of which were then transformed to *Z*-scores. The variability of these correlations was calculated as their standard deviation, and subjects with outlying correlation variability values (±3 standard deviations from the mean) were removed. We also calculated the static functional connectivity strength between these regions using the entire, non-windowed time series to understand the relationship between static and dynamic functional connectivity. Further statistical analyses were performed using R.[27] All the group comparisons and correlations were adjusted for age, sex and head motion (total distance travelled), and were corrected for the number of comparisons (mPFC–PCC and two negative controls) using false discovery rate correction.

We assessed relationships between functional connectivity variability and clinical measures including depression and anxiety severity scores, time since illness onset and RRS in Pearson partial correlations.

Voxel-based morphometry

We also compared grey matter volumes in the chosen ROIs between patients and controls using voxel-based morphometry to examine the co-existence of volumetric changes that may explain changes in connectivity. High-resolution T1-weighted structural images (repetition time = 7.31 ms, time to echo = 3.02 ms, 256 × 256 matrix, 196 slices, voxel size = 1.2 × 1.05 × 1.05 mm, for both the samples) were pre-processed using voxel-based morphometry in SPM12 (*www.fil.ion.ucl.ac.uk/spm*) for both the samples. Images were segmented into different tissue types and processed with DARTEL[28] before being normalized to Montreal Neurological Institute space. The modulated grey matter images were then smoothed with an 8 mm full width at half maximum Gaussian kernel. Grey matter volume was next compared between groups within the same mPFC and PCC regions of interest used in the functional connectivity analysis with a two-sample *t*-test. A cluster-defining voxelwise threshold of $P < 0.001$ uncorrected was used, with a clusterwise threshold of 0.05 false discovery rate corrected. Total grey matter volume was also calculated based on the segmented maps created in SPM (thresholded at grey matter volume probability >0.5) and compared between the groups.

RESULTS

Participants

Twenty patients with unipolar major depression in the primary study and 19 in the validation study were sex and age matched with 19 healthy controls in the primary sample and 20 in the validation sample (Table 1). Three and nine subjects were recruited through psychological therapy services in the primary and validation samples, respectively, while the remaining participants were recruited from the community. A healthy participant from the validation sample was removed from the analyses due to outlying connectivity variability values. Figure 1 shows coordinates for the selected ROIs.

Connectivity variability in major depression

Connectivity variability, expressed as the standard deviation of connectivity strength between the mPFC and the PCC (Figure 2), was significantly greater in patients with major depression versus healthy controls ($t(37) = 2.56$, $P = 0.044$, $d = 0.82$). This effect was successfully replicated in the validation sample ($t(36) = 2.53$, $P = 0.045$, $d = 0.82$) supporting the coherence of the model in identifying consistently greater connectivity variability across samples irrespective of clinical heterogeneity.

There were no group differences in connectivity variability in either sample between the mPFC and primary motor cortex, chosen as a negative control region (primary sample: $t(37) = 0.79$, $P = 0.44$, validation sample: $t(36) = 1.85$, $P = 0.22$), or between the PCC and primary motor cortex (primary sample: $t(37) = 1.76$, $P = 0.17$, validation sample: PCC: $t(36) = 0.63$, $P = 0.99$), suggesting that results did not reflect global instability. No differences in static connectivity strength between mPFC and PCC were found between patients and controls in either sample (primary sample: $t(37) = 0.33$, $P = 0.74$, $d = 0.11$, validation sample: $t(36) = 0.73$, $P = 0.47$, $d = 0.24$).

Connectivity variability and cognitive style

A positive correlation of RRS with connectivity variability was noted in the validation sample ($r(14) = 0.51$, $P = 0.045$, Figure 3), and not the primary sample ($r(15) = 0.18$, $P = 0.48$). The relationship did not remain significant in a pooled analysis with both groups combined and sample as a covariate ($r(34) = 0.30$, $P = 0.075$). We did not find significant correlations between RRS scores and static connectivity strength in either sample (primary sample: $r(15) = 0.31$, $P = 0.22$, validation sample: $r(14) = -0.13$, $P = 0.62$).

Table 1. Samples clinical characteristics

	Major depression	*Healthy control*	P
Primary sample			
n	20	19	
Age, years	29.55 (6.59)	30.05 (6.71)	0.81
Male/female	2, 18	2, 18	1
MÅDRS	27.25 (4.24)	0.95 (1.39)	< 0.001
HDRS anxiety subscale	4.95 (2.48)	0.21 (0.42)	< 0.001
RRS	56.00 (10.23)	29.67 (6.44)	< 0.001
Time since illness onset, years	6.35 (6.41)	—	—
Comorbid diagnoses	None	—	—
Hospitalizations	0	—	—
Number of episodes	1.5 (1.25)	—	—
Validation sample			
n	19	19	
Age, years	32.34 (10.62)	31.91 (10.30)	0.90
Male/female	7, 12	6, 13[a]	0.73
MÅDRS	30.74 (7.31)	1.37 (1.86)	< 0.001
HDRS anxiety subscale	7.16 (1.30)	0.31 (0.58)	< 0.001
RRS	66.47 (8.22)	30.63 (6.83)	< 0.001
Time since illness onset, years	13.50 (8.26)	—	—
Comorbid diagnoses	*N* = 12 (9 GAD, 5 SAD, 4 OCD, 2 PD, 2 PTSD, 1 historic substance abuse)	—	—
Hospitalizations	4	—	—
Number of episodes	4 (2.5)	—	—

Abbreviations: GAD, generalized anxiety disorder; HDRS, Hamilton Depression Rating Scale (17 item); MÅDRS, Montgomery-Åsberg Depression Rating Scale; OCD, obsessive compulsive disorder; PD, panic disorder; PTSD, post-traumatic stress disorder; RRS, Ruminative Response Scale; SAD, social anxiety disorder. [a]Demographics for the validation sample represent the 19 healthy controls included in the final analysis. The samples did not differ on depression severity ($t(37) = -1.83$, $P = 0.07$). The validation sample had a significantly higher anxiety score ($t(37) = 3.45$, $P = 0.001$), time since illness onset ($t(37) = 2.62$, $P = 0.01$) and RRS score ($t(37) = 3.51$, $P = 0.001$) than the primary sample. Values are reported as mean (standard deviations) for all variables except number of episodes where due to skewed data we report median (interquartile range).

Correlations with clinical variables

Correlations between connectivity variability and depressive symptom severity were not significant (primary sample: $r(15) = 0.37$, $P = 0.14$, validation sample: $r(14) = 0.23$, $P = 0.40$). The same pattern of results was observed between correlation variability and anxiety symptoms scores, as measured by the anxiety subscale of the Hamilton Depression Rating Scale (sample A: $r(15) = 0.11$, $P = 0.66$, sample B: $r(14) = -0.23$, $P = 0.39$). The time since illness onset was not significantly correlated with connectivity variability in either sample (sample A: $r(15) = 0.25$, $P = 0.33$, sample B: $r(14) = 0.16$, $P = 0.55$).

Grey matter volumes

There were no differences between groups in grey matter volume in the chosen regions of interest (no significant clusters at $P < 0.001$, false discovery rate corrected). There were also no differences in the total grey matter volume between groups (primary sample: $t(37) = 1.43$, $P = 0.16$, validation sample: $t(36) = 0.79$, $P = 0.43$).

Head motion

There were no significant differences in the total distance travelled between the patients and controls in the primary sample

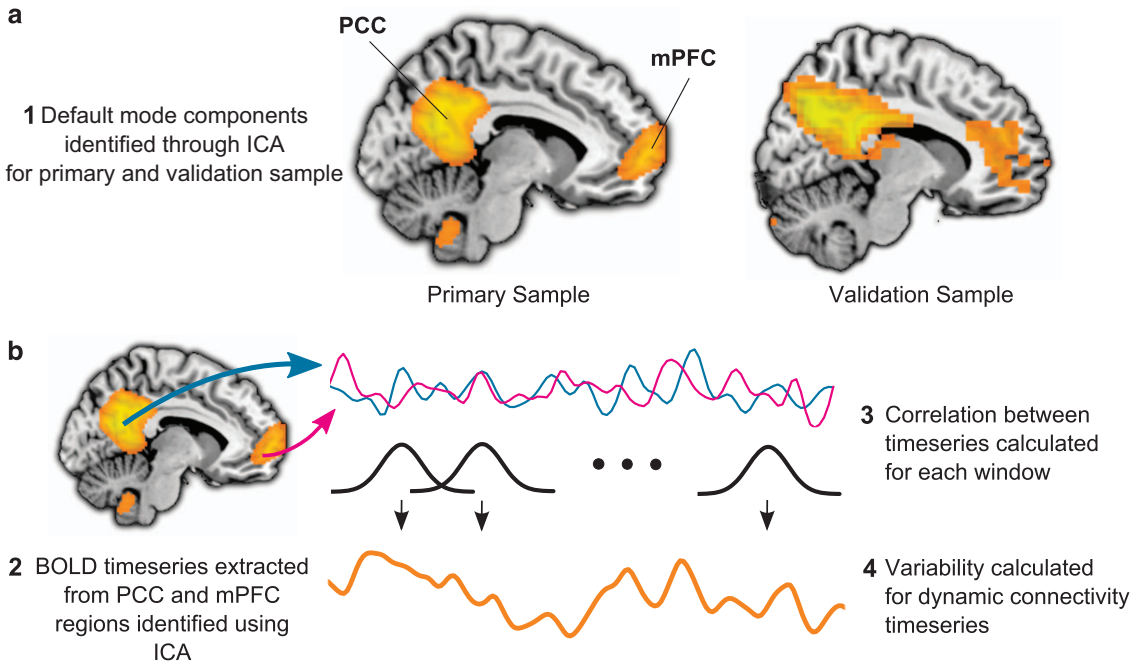


Figure 1. Illustration of the dynamic functional connectivity analysis method. In the primary sample, sample-specific Montreal Neurological Institute (MNI) coordinates for the posterior cingulate cortex (PCC) were 2, −62, 22 and for the medial prefrontal cortex (mPFC) 4, 60, 0. In the validation sample, MNI coordinates were 6, −44, 11 and 2, 60, −4, respectively. (**a**) Default mode network (DMN) components for each sample identified using group independent component analysis (ICA), showing clusters in the mPFC and PCC. (**b**) Illustration of the dynamic functional connectivity method. BOLD, blood oxygen level dependent.

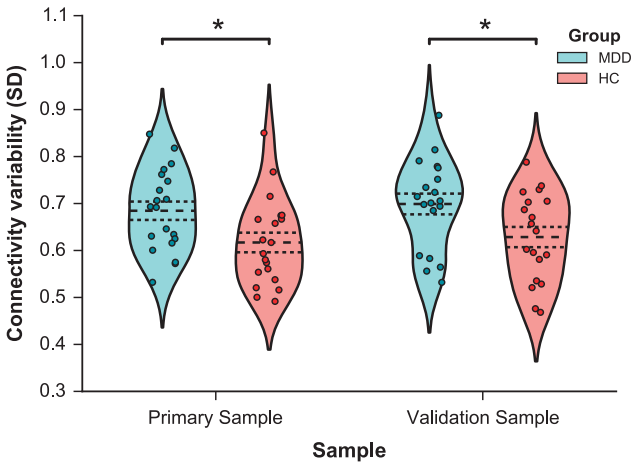


Figure 2. Variability in connectivity strength for patient and healthy control groups in both samples. Plots represent the distribution of data for each group, along with individual data points. Dashed lines represent means and standard errors. *$P < 0.05$ ($P = 0.044$ for the primary sample and $P = 0.048$ for the validation sample). HC, healthy controls; MDD, major depressive disorder.

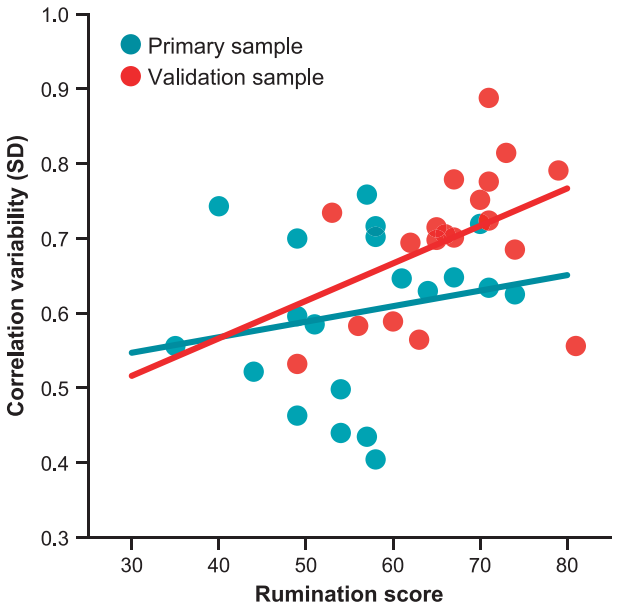


Figure 3. Correlation between connectivity variability and ruminative response style (RRS) score in major depression in the primary and validation sample. RRS, Rumination Response Scale.

($t(37) = -1.31$, $P = 0.20$) or the validation sample ($t(36) = 0.56$, $P = 0.58$). When looking at framewise motion measures, there were no significant differences between the groups in mean framewise displacement (primary sample: $t(37) = 0.88$, $P = 0.38$, validation sample: $t(36) = 0.26$, $P = 0.80$) or maximum displacement (primary sample: $t(37) = 0.42$, $P = 0.68$, validation sample: $t(36) = 1.48$, $P = 0.15$). There was also no significant difference between groups in the number of interpolated time points (primary sample: $U(37) = 173$, $P = 0.64$, non-parametric test used due to non-normally distributed data, validation sample: $t(36) = 1.60$, $P = 0.12$), or in the thresholds used for detecting outlying time points for either the primary sample (REFRMS: $t(37) = -0.14$, $P = 0.89$, DVARS: $t(37) = 0.42$, $P = 0.68$) or the validation sample (REFRMS: $t(36) = 0.67$, $P = 0.51$, DVARS: $t(36) = 0.71$, $P = 0.48$).

There were also no significant correlations in the validation sample between rumination scores and total distance travelled ($R(17) = 0.13$, $P = 0.59$), mean displacement ($R(17) = -0.09$, $P = 0.70$), maximum displacement ($R(17) = 0.17$, $P = 0.48$), or the number of interpolated time points ($R(17) = 0.11$, $P = 0.66$), indicating that the observed relationship between connectivity variability cannot be explained by motion. There were no significant relationships between thresholds used to detect outlying time points and rumination scores (REFRMS $R(17) = 0.11$, $P = 0.65$, DVARS: $R(17) = 0.35$, $P = 0.15$).

DISCUSSION

We compared variability in connectivity strength within the DMN between medication-free individuals with major depression and matched healthy controls. We found that connectivity between the mPFC and PCC, two key nodes in the DMN, was significantly more variable in currently symptomatic patients with major depression. Furthermore, the validity of the results was confirmed in an independent sample of individuals with significantly more clinical heterogeneity, and using multi-echo acquisition parameters to limit the impact of non-neural signals. We believe this is the first time that greater connectivity variability has been reported in major depression in this DMN subsystem, complementing findings of reported abnormal dynamic connectivity in other brain regions in major depression[13] and other psychiatric conditions.[11,29,30]

In agreement with Kaiser *et al*.,[13] we demonstrated that the increased variability in neural connectivity originating from the mPFC correlates with a ruminative thinking pattern, indicating a possible synergy of this DMN subsystem with the 'salience network' represented by the insula.[14] This previous study, a whole-brain analysis, demonstrated altered connectivity variability between the mPFC and other DMN regions, but not specifically the PCC,[13] and variability in connectivity between the mPFC and parahippocampal gyrus was reduced. This might be explained by the more limited power in whole-brain analyses due to the necessary correction for multiple comparisons. Another possible explanation is that depression is associated with increased connectivity variability in the central DMN (including the PCC), while variability in the ventral system (including the parahippocampal gyrus) is reduced.

At present, the biological significance of time varying properties of connectivity is not well understood.[12] One potential explanation for the observed greater connectivity variability is that it results from reduced structural connectivity in the DMN as suggested by previous research linking structural integrity with connectivity variability.[31] However, our recent meta-analysis reported that structural integrity is maintained in the tracts connecting these regions in major depression,[7] suggesting that such an explanation is unlikely, further supported by our finding of no alteration in grey matter volume in the ROIs in these patient samples. It is also possible that greater connectivity variability reflects primary alterations in neuronal communication rather than occurring secondary to aberrant structural connectivity, a suggestion that is in line with preclinical work showing primary abnormalities of neural processing in circuits relevant to depression.[32] Clarifying the precise meaning of connectivity variability, and how it relates to connectivity strength, at a neural level will be an important task for future research.

We also tested the hypothesis that a ruminative cognitive style, linked with depression,[8,33] potentially explains the observed abnormalities in connectivity variability between mPFC and PCC. We observed a direct positive correlation between levels of intrusive, self-generated, ruminative thoughts and variability in connectivity within this network. This was, however, present only in the validation sample. This discrepancy in the results can be explained by issues related to statistical power combined with variability in the samples clinical characteristics. Patients in the validation sample were in fact characterized by higher rumination scores and comorbid anxiety disorders. It is possible that this result is reflective of a stronger link between anxiety symptoms and rumination,[34] and further work could clarify this. The difference between samples may also be related to differences in image acquisition and preprocessing. It is also possible that connectivity variability might be more closely linked to 'state' rumination levels occurring at the time of scanning rather than more ingrained 'trait' measures such as the RRS or to state anxiety, which we did not assess here. This is supported by work adopting post-scan cognitive style reports in healthy volunteers.[35] This study suggested a positive correlation between variability in the DMN and reported daydreaming (a related phenomenon to rumination) during MRI scanning, consistent with our findings.

We found no association in either sample between variability and depression severity scores. This may suggest that increased variability is not directly related to depressive symptoms. However, it is possible that this could be due to the Montgomery–Åsberg Depression Rating Scale, our measure of symptom severity, being weighted towards somatic rather than cognitive symptoms of depression. Previous research suggests that alterations in DMN connectivity are more likely to be related to psychological symptoms such as negative self-related cognitions.[33] Another possible explanation is that the relationship between biological disease mechanisms and symptoms is complex, and tend not to correlate linearly with one another.[11]

We found no alterations in static connectivity strength in both our samples in agreement with some studies[36] but not others.[37] Discrepancies in the findings can be explained by methodological differences in connectivity measurements, the proneness of neuroimaging data to type I/II error based on their relatively moderate sample sizes as indicated by recent meta-analyses[4] and multiple sources of heterogeneity intrinsic to major depressive disorders. We adopted a cross-validation method applied for the first time to this type of data to help enhance robustness of the findings. Based on our experience, future studies could consider a similar approach with larger samples, perhaps in the context of collaborative mega-analyses[38] to reduce statistical bias and increase power.

In addition to this potential link between connectivity instability and alterations in cognitive or emotional state, such as rumination, instability may also reflect increased noise or alterations in neural dynamics. Simulation studies have indicated that patterns of synchronization and desynchronization in neuronal populations lead to fluctuations in functional connectivity as measured using fMRI,[39] while noise-driven neuronal simulations produce switches between states of functional connectivity.[40] Notably, simultaneous electroencephalography and fMRI have shown that changes in the BOLD functional connectivity mirror electroencephalography power fluctuations[41] further indicating that variations in the functional connectivity are reflective of neuronal processes. It is possible that fluctuations in connectivity may reflect underlying changes in neural synchrony, a key process in inter-regional communication, which has been proposed to be affected in a range of psychiatric disorders.[42]

It has previously been suggested that hyperconnectivity implies lower variability in connectivity.[13,43] However, this is at odds with our results as we did not find group differences in static connectivity. Hence, the relationship between dynamic and static connectivity appears complex, with dynamic functional connectivity providing distinct information about network communication in a state of pathology independent from and beyond that of static connectivity. This echoes findings from previous research that attempted to classify patients with schizophrenia, bipolar disorder and healthy controls based on functional connectivity, indicating that classification using a combination of static and dynamic connectivity tends to be more accurate than static connectivity alone.[11,44]

A notable strength of this study is the control of non-neural influences on the data. The resting-state fMRI analyses are susceptible to influence from confounding factors such as motion and physiological variables,[24] and this is especially pertinent here in view of the possibility that physiological characteristics such as heart rate may differ between groups due to increased anxiety in patient groups. Consequently, we rigorously controlled for these in a number of ways, including correction for cardiac and respiratory signals, and motion scrubbing.[24] In our replication sample we used multi-echo fMRI with independent component analysis-based de-noising, which is more effective than traditional de-noising methods,[15] providing further evidence that our original results were not a result of non-neural influences. Moreover, our negative control analyses indicate that our findings were specific to the disease-relevant network under investigation, and groups did not differ on motion parameters.

Further strengths of the present study are the inclusion of medication-free patients, suggesting that the findings are not due to pharmacological effects. In addition, many patients in the primary sample were medication-naive and had experienced few, if any, past depressive episodes, making it less likely that the effects observed are cumulative effects of illness or previous treatment. The demonstration of greater connectivity variability in the validation sample, which included patients with more chronic and heterogeneous illness and used multi-echo fMRI, increases confidence in the reliability of the findings, which are unlikely to be attributable to idiosyncrasies of a specific sample or methodological artefacts. This is particularly important in view of concerns regarding poor reproducibility in research.[45] Replication is an important step towards minimizing false associations and enhancing reliability of results,[45] and the reproducibility demonstrated here suggests that alterations in dynamic functional connectivity are robust. Future replications of these results are warranted, and it would be of particular interest to examine connectivity variability in non-symptomatic individuals with depression and 'at risk' samples to test whether it is a feature of the depressive state, a marker of vulnerability to depression or a 'scar' effect.

One limitation of this study is the focus on two isolated regions of the DMN. We chose these regions given their key role in the DMN, and association with rumination in major depression.[8,9] Focusing on *a priori* regions of interest nevertheless increased the power to detect changes given the relatively small sample sizes. Additional analyses to confirm the replicability of previous findings, such as altered dynamic connectivity with the insula, would have been of interest; however, we chose to focus on one particular component of the DMN to reduce the likelihood of type II errors due correction for multiple comparisons with a small sample.

Furthermore, alternative methods to sliding window analyses, such as coherence-based methods, have been proposed that may provide more accurate estimations of dynamic functional connectivity.[12] Given their novelty, we have however chosen to use a more established method that has been evaluated in numerous studies. It is also possible that using different window lengths may affect the analysis. We believe the 40 s windows used here is an optimal length for detecting alterations in dynamic functional connectivity (see Supplementary Material for further discussion of this issue). In addition, our imaging sequences were only 6 and 8 min long for the primary and replication sample respectively. This may have limited our ability to detect less-frequent fluctuations in connectivity. Furthermore, our replication data and analysis differed slightly from our primary sample in acquisition and de-noising methods, making this more of a conceptual than methodological replication. Nevertheless, the fact that our results were largely consistent despite these dissimilarities suggests that the effect is robust.

In view of the presence of a range of anxiety disorders in the validation sample, it is not possible to determine whether altered connectivity variability might be a common abnormality present in both major depression and anxiety disorders rather than being specifically associated with depression. High comorbidity rates between anxiety and depressive disorders and the co-occurrence of anxiety symptoms in major depression are frequent findings in clinical practice.[1] This makes the differentiation challenging to establish. Lastly, it should be noted that the sample sizes used here were relatively small. Nonetheless, the consistency of our results across two independent samples indicates that our results are unlikely to be spurious.

In conclusion, our study indicates that major depression is associated with reduced stability of connectivity within the DMN in key regions relevant to the generation of ruminative cognitions. This could represent an intrinsic neural property of this illness and a potential network-specific brain abnormality not previously explained by structural abnormalities or static functional connectivity. The results were replicated in a second independent sample, indicating that they are robust. Ruminative cognitive style might partially explain the results in keeping with cognitive models of depressive disorders. This work adds to functional connectivity research in affective disorders by validating new findings across different samples with novel fMRI analysis techniques. Further work investigating trait markers of vulnerability to depression and connectivity variability in remitted patients are necessary to establish the specificity of these findings to the depressed state. Similarly, longitudinal studies are required to explore the effect of treatment on these abnormalities.

CONFLICT OF INTEREST

AJC has in the last 3 years received honoraria for speaking from AstraZeneca, honoraria for consulting from Allergan and Livanova and research grant support from Lundbeck. DJL has consulted for Ixico PLC. AHY has given paid lectures and sits on advisory boards for all major pharmaceutical companies with drugs used in affective and related disorders; is the lead investigator for the Embolden Study (AZ), BCI Neuroplasticity study and Aripiprazole Mania Study and investigator-initiated studies from AstraZeneca, Eli Lilly, Lundbeck and Wyeth; and has received grant funding from the National Institute of Mental Health (USA); Canadian Institutes of Health Research (Canada); National Alliance for Research on Schizophrenia and Depression (USA); Stanley Medical Research Institute (USA); Medical Research Council (UK); Wellcome Trust (UK); Royal College of Physicians (Edinburgh); British Medical Association (UK); University of British Columbia-Vancouver General Hospital Foundation (Canada); Western Economic Diversification Canada (Canada); Canadian Cardiovascular Society Depression Research Fund (Canada); Michael Smith Foundation for Health Research (Canada); and National Institute for Health Research (UK). AMP is supported by Bionomics Limited. DA has received travel grants from Janssen-Cilag and Servier. SCRW has received grant funding from the Medical Research Council (UK); Wellcome Trust (UK); National Institute for Health Research (UK) and support for investigator-led studies from Takeda, Pfizer, Lundbeck, P1Vital, Roche and Eli Lilly. The remaining authors declare no conflict of interest.

ACKNOWLEDGMENTS

The primary study was funded by Academy of Medical Sciences grant AMS-SGCL8 to DA and a National Institute of Health Research (NIHR) PhD studentship to TW supervised by AJC and DA. The validation study was funded by a Medical Research Council (MRC)/IoPPN Excellence PhD studentship to LM, supervised by AJC and AMP, and departmental funds generated by AJC and SCRW. AJC, AMP, SCRW, TW and AHY are supported by the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. This study represents independent research part funded by the NIHR/Wellcome Trust, King's College Hospital Clinical Research Facility and the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. AH-V was supported by a Chilean Bicentennial Fund Scholarship from the Bicentennial Fund for Human Capital Development (Becas Chile) and by the Psychiatric Research Trust. We thank the staff of the NIHR/Wellcome Trust Clinical Research Facility at King's College Hospital, the Centre for Neuroimaging Science at King's College London and Dean Broadhurst, and Michael Kelly of the MHRN for their support in the conduct of the study. We thank all the participants of this study for their support.

DISCLAIMER

The funders had no role in the design and conduct of the study, collection, management, analysis and interpretation of the data, preparation, review, approval of the manuscript and decision to submit the manuscript for publication. The views expressed are those of the author(s) and not necessarily those of the NHS, NIHR, MRC, Mental Health Research Network (MHRN), the Academy of Medical Sciences, or the Department of Health.

REFERENCES

1 Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR *et al.* The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *J Am Med Assoc* 2003; **289**: 3095–3105.
2 Üstün TB, Ayuso-Mateos JL, Chatterji S, Mathers C, Murray CJL. Global burden of depressive disorders in the year 2000. *Br J Psychiatry* 2004; **184**: 386–392.
3 Wise T, Cleare AJ, Herane A, Young AH, Arnone D. Diagnostic and therapeutic utility of neuroimaging in depression: an overview. *Neuropsychiatr Dis Treat* 2014; **10**: 1509–1522.
4 Kaiser RH, Andrews-Hanna JR, Wager TD, Pizzagalli DA. Large-scale network dysfunction in major depressive disorder: a meta-analysis of resting-state functional connectivity. *JAMA Psychiatry* 2015; **72**: 603–611.
5 Arnone D, Job D, Selvaraj S, Abe O, Amico F, Cheng Y *et al.* Computational meta-analysis of statistical parametric maps in major depression. *Hum Brain Mapp* 2016; **37**: 1393–1404.
6 Wise T, Radua J, Via E, Cardoner N, Abe O, Adams T *et al.* Common and distinct patterns of grey matter volume alteration in major depression and bipolar disorder: evidence from voxel-based meta-analysis. *Mol Psychiatry* 2016.
7 Wise T, Radua J, Nortje G, Cleare AJ, Young AH, Arnone D. Voxel-based meta-analytical evidence of structural disconnectivity in major depression and bipolar disorder. *Biol Psychiatry* 2016; **79**: 293–302.
8 Perkins AM, Arnone D, Smallwood J, Mobbs D. Thinking too much: self-generated thought as the engine of neuroticism. *Trends Cogn Sci* 2015; **19**: 492–498.
9 Andrews-Hanna JR, Reidler JS, Sepulcre J, Poulin R, Buckner RL. Functional-anatomic fractionation of the brain's default network. *Neuron* 2010; **65**: 550–562.
10 Berman MG, Misic B, Buschkuehl M, Kross E, Deldin PJ, Peltier S *et al.* Does resting-state connectivity reflect depressive rumination? A tale of two analyses. *Neuroimage* 2014; **103**: 267–279.
11 Calhoun VD, Miller R, Pearlson G, Adalı T. The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron* 2014; **84**: 262–274.
12 Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M *et al.* Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* 2013; **80**: 360–378.
13 Kaiser RH, Whitfield-Gabrieli S, Dillon DG, Goer F, Beltzer M, Minkel J *et al.* Dynamic resting-state functional connectivity in major depression. *Neuropsychopharmacology* 2016; **41**: 1822–1830.
14 Uddin LQ. Salience processing and insular cortical function and dysfunction. *Nat Rev Neurosci* 2015; **16**: 55–61.
15 Kundu P, Inati SJ, Evans JW, Luh W-M, Bandettini PA. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *Neuroimage* 2012; **60**: 1759–1770.
16 Wise T, Arnone D, Marwood L, Zahn R, Lythe K, Young A. Recruiting for research studies using online public advertisements: examples from research in affective disorders. *Neuropsychiatr Dis Treat* 2016; **12**: 279–285.
17 Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998; **59**: 22–33-57.
18 Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979; **134**: 382–389.
19 Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; **23**: 56–62.
20 McClintock SM, Husain MM, Bernstein IH, Wisniewski SR, Trivedi MH, Morris D *et al.* Assessing anxious features in depressed outpatients. *Int J Methods Psychiatr Res* 2011; **20**: e69–e82.
21 Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971; **9**: 97–113.
22 Treynor W, Gonzalez R, Nolen-Hoeksema S. Rumination reconsidered: a psychometric analysis. *Cogn Ther Res* 2003; **27**: 247–259.
23 Glover GH, Li T-Q, Ress D. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn Reson Med* 2000; **44**: 162–167.
24 Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 2012; **59**: 2142–2154.
25 Varoquaux G, Sadaghiani S, Pinel P, Kleinschmidt A, Poline JB, Thirion B. A group model for stable multi-subject ICA on fMRI datasets. *Neuroimage* 2010; **51**: 288–299.
26 Zalesky A, Breakspear M. Towards a statistical test for functional connectivity dynamics. *Neuroimage* 2015; **114**: 466–470.
27 R Core Team. *R: A Language and Environment for Statistical Computing [Internet]*. R Foundation for Statistical Computing: Vienna, Austria, 2015. Available fromhttps://www.R-project.org/.
28 Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage* 2007; **38**: 95–113.
29 Kopell NJ, Gritton HJ, Whittington MA, Kramer MA. Beyond the connectome: the dynome. *Neuron* 2014; **83**: 1319–1328.
30 Damaraju E, Allen EA, Belger A, Ford JM, McEwen S, Mathalon DH *et al.* Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *Neuroimage Clin* 2014; **5**: 298–308.
31 Shen K, Hutchison RM, Bezgin G, Everling S, McIntosh AR. Network structure shapes spontaneous functional connectivity dynamics. *J Neurosci* 2015; **35**: 5579–5588.
32 Russo SJ, Nestler EJ. The brain reward circuitry in mood disorders. *Nat Rev Neurosci* 2013; **14**: 609–625.
33 Hamilton JP, Furman DJ, Chang C, Thomason ME, Dennis E, Gotlib IH. Default-mode and task-positive network activity in major depressive disorder: implications for adaptive and maladaptive rumination. *Biol Psychiatry* 2011; **70**: 327–333.
34 McLaughlin KA, Nolen-Hoeksema S. Rumination as a transdiagnostic factor in depression and anxiety. *Behav Res Ther* 2011; **49**: 186–193.
35 Kucyi A, Davis KD. Dynamic functional connectivity of the default mode network tracks daydreaming. *Neuroimage* 2014; **100**: 471–480.
36 Sheline YI, Price JL, Yan Z, Mintun MA. Resting-state functional MRI in depression unmasks increased connectivity between networks via the dorsal nexus. *Proc Natl Acad Sci USA* 2010; **107**: 11020–11025.
37 Greicius MD, Flores BH, Menon V, Glover GH, Solvason HB, Kenna H *et al.* Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biol Psychiatry* 2007; **62**: 429–437.
38 Hallahan B, Newell J, Soares JC, Brambilla P, Strakowski SM, Fleck DE *et al.* Structural magnetic resonance imaging in bipolar disorder: an international collaborative mega-analysis of individual adult patient data. *Biol Psychiatry* 2011; **69**: 326–335.
39 Honey CJ, Kötter R, Breakspear M, Sporns O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc Natl Acad Sci USA* 2007; **104**: 10240–10245.
40 Hansen ECA, Battaglia D, Spiegler A, Deco G, Jirsa VK. Functional connectivity dynamics: modeling the switching behavior of the resting state. *Neuroimage* 2015; **105**: 525–535.
41 Tagliazucchi E, von Wegner F, Morzelewski A, Brodbeck V, Laufs H. Dynamic BOLD functional connectivity in humans and its electrophysiological correlates. *Front Hum Neurosci* 2012; **6**: 339.
42 Voytek B, Knight RT. Dynamic network communication as a unifying neural basis for cognition, development, aging, and disease. *Biol Psychiatry* 2015; **77**: 1089–1097.
43 Chang C, Glover GH. Time–frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage* 2010; **50**: 81–98.
44 Rashid B, Damaraju E, Pearlson GD, Calhoun VD. Dynamic connectivity states estimated from resting fMRI identify differences among schizophrenia, bipolar disorder, and healthy control subjects. *Front Hum Neurosci* 2014; **8**: 897.
45 Ioannidis JPA. Why science is not necessarily self-correcting. *Perspect Psychol Sci* 2012; **7**: 645–654.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2017

Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)

Report

Current Biology

Intermittent Stimulation of the Nucleus Basalis of Meynert Improves Working Memory in Adult Monkeys

Highlights

- Intermittent stimulation of nucleus basalis (NB) improves working memory performance
- Continuous NB stimulation degrades performance in young adult monkeys
- Donepezil provided no greater improvement than intermittent stimulation
- Effects were blocked by nicotinic and muscarinic receptor antagonists

Authors

Ruifeng Liu, Jonathan Crawford, Patrick M. Callahan, Alvin V. Terry, Jr., Christos Constantinidis, David T. Blake

Correspondence

dblake.mcg@gmail.com

In Brief

Liu et al. show that an intermittent train of electrical stimulation in the nucleus basalis improves working memory in adult monkeys. The improvement depends on cholinergic receptors. The finding reveals a new mechanism to improve cognitive performance in conditions that compromise cholinergic action, including aging and Alzheimer's disease.

Liu et al., 2017, Current Biology *27*, 1–7
September 11, 2017 © 2017 Elsevier Ltd.
http://dx.doi.org/10.1016/j.cub.2017.07.021

CellPress

Please cite this article in press as: Liu et al., Intermittent Stimulation of the Nucleus Basalis of Meynert Improves Working Memory in Adult Monkeys, Current Biology (2017), http://dx.doi.org/10.1016/j.cub.2017.07.021

Current Biology
Report

Intermittent Stimulation of the Nucleus Basalis of Meynert Improves Working Memory in Adult Monkeys

Ruifeng Liu,[1,2] Jonathan Crawford,[2] Patrick M. Callahan,[3] Alvin V. Terry, Jr.,[3] Christos Constantinidis,[4] and David T. Blake[2,5,*]
[1]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China
[2]Brain and Behavior Discovery Institute, Department of Neurology, Medical College of Georgia, Augusta University, 1120 15th Street, Augusta, GA 30912, USA
[3]Department of Pharmacology and Toxicology, Medical College of Georgia, Augusta University, 1120 15th Street, Augusta, GA 30912, USA
[4]Department of Neurobiology and Anatomy, Wake Forest School of Medicine, Winston-Salem, NC, USA
[5]Lead Contact
*Correspondence: dblake.mcg@gmail.com
http://dx.doi.org/10.1016/j.cub.2017.07.021

SUMMARY

Acetylcholine in the neocortex is critical for executive function [1–3]. Degeneration of cholinergic neurons in aging and Alzheimer's dementia is commonly treated with cholinesterase inhibitors [4–7]; however, these are modestly effective and are associated with side effects that preclude effective dosing in many patients [8]. Electrical activation of the nucleus basalis (NB) of Meynert, the source of neocortical acetylcholine [9, 10], provides a potential method of improving cholinergic activation [11, 12]. Here we tested whether NB stimulation would improve performance of a working memory task in a nonhuman primate model. Unexpectedly, intermittent stimulation proved to be most beneficial (60 pulses per second, for 20 s every minute), whereas continuous stimulation often impaired performance. Pharmacological experiments confirmed that the effects depended on cholinergic activation. Donepezil, a cholinesterase inhibitor, restored performance in animals impaired by continuous stimulation but did not improve performance further during intermittent stimulation. Intermittent stimulation was rendered ineffective by either nicotinic or muscarinic receptor antagonists. In the months after stimulation began, performance also improved in sessions without stimulation. Our results reveal that intermittent NB stimulation can improve working memory, a finding that has implications for restoring cognitive function in aging and Alzheimer's dementia.

RESULTS AND DISCUSSION

Five adult Rhesus monkeys were trained with a delayed match-to-sample task, requiring them to remember a stimulus presented on a touch screen and, after a delay period, to select the stimulus that was initially presented (Figure 1). Once the animals were proficient with the task, they were implanted bilaterally with NB stimulation electrodes. Placement and anatomical verification are described in STAR Methods. The zone of stimulation was broad, on the order of a 4-mm-diameter sphere [15], and the accuracy of placement of the electrode tip within the nucleus was not critical in this experiment. Experiments in the first two animals (K and S) primarily guided proper positioning of stimulation leads in the next three (CH, DI, and PU). No histological confirmation of electrode placement occurred in these three animals, and electrode positioning is presumptive based on stereotaxic targeting.

Effects of Electrical Stimulation Parameters

We initially tested the hypothesis that NB stimulation improves working memory performance by applying a continuous train of stimulation pulses in blocks of 100 trials interleaved with blocks of 100 trials without stimulation. Contrary to our expectations, we found that continuous stimulation always impaired performance, and effects were larger at higher stimulation rates (Figure 2A). Results for continuous stimulation at 80 Hz reached statistical significance (binomial tests; animal CH: $n = 800$ trials, $p < 0.0001$; animal DI: $n = 1000$, $p < 0.001$).

In an effort to determine whether stimulation during a particular interval of the task was disrupting working memory, the task was altered to stimulate only during the inter-trial period or to stimulate only during trials. Unexpectedly, either condition resulted in supranormal performance. We established an intermittent stimulation condition that provided 20 s of pulses at a rate of 60 per second followed by 40 s without pulses. Results from this condition are shown in Figure 2B. Trials with stimulation resulted in better performance than trials with no stimulation in the animals tested (binomial test, $n = 1000$, $p < 0.01$). Note that a longer delay period was used in Figure 2B than in Figure 2A to avoid ceiling effects. We conclude that intermittent stimulation results in supranormal working memory performance.

We explored the parameters for optimal stimulation in two ways. First, the number of pulses per minute was fixed, and the duration and rate of stimulation were altered. The 1,200 pulses per minute were delivered in 10 s (i.e., at a rate of

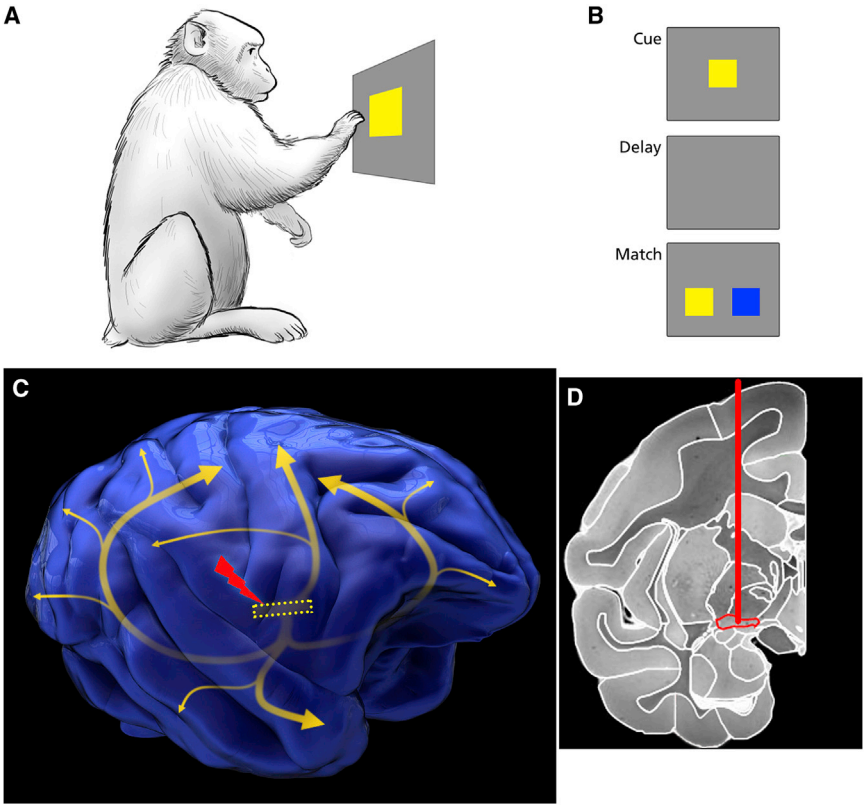


Figure 1. Task Paradigm and Electrical Stimulation Position in the Brain

(A) Macaque monkeys were trained to interact with a touchscreen. The first step in the task is touching the cue.

(B) After the cue is touched, the screen blanks during the delay period, followed by presentation of two potential matches. Three colors were used in training, and two were randomly selected as cue/match and distractor on each trial.

(C) Stimulation of the nucleus basalis (NB) of Meynert was used in experimental sessions. The position of the nucleus is shown by the dotted yellow rectangle, with the curved arrows approximating the pathways from the nucleus to neocortices [13].

(D) An MRI coronal section of a Rhesus monkey brain with the implantation target outlined in red. The MRI was taken from the Macaque Scalable Brain Atlas [14] (M.F. Dubach, and D.M. Bowden, 2009, Soc. Neurosci., abstract).

See also Figures S1 and S2.

120 Hz), 20 s, 40 s, or 60 s. Results are shown in Figure 2C for animals CH and DI. A two-way ANOVA found a significant effect of these conditions on performance (test for animal number and condition, $F(4,4) = 12.92$ for main effect of condition, $p < 0.02$; $F(1,4) = 1.48$ for main effect of animal, $p > 0.2$). Post hoc comparisons showed that performance was significantly better at the 10 or 20 s period than the 60 s condition of continuous stimulation (binomial statistic, ~1,050 trials for each condition, Bonferroni-corrected $p < 0.005$).

In a second experiment, the rate of stimulation was fixed at 80 pulses per second but the duration was varied to 10 s, 20 s, 40 s, or 60 s per minute (Figure 2D). A two-way ANOVA found a significant effect of condition and animal ($F(1,4) = 15.09$ for animal, $p < 0.02$; $F(4,4) = 8.48$ for condition, $p < 0.04$). The highest performance occurred for the 20 s stimulation condition, which was significantly better than the 60 s stimulation in post hoc tests (binomial statistic, ~1,350 trials each, Bonferroni-corrected $p < 0.005$). The significant ANOVA finding on the animal variable occurred as animal CH averaged higher performance in the 80 Hz stimulation conditions than animal DI. These parametric tests revealed that the condition of 1,200 pulses being delivered 15–20 s per minute was optimal within the parameters tested.

The effect of intermittent stimulation across varying delay periods can be seen in Figures 2E and 2F. The performance change, expressed as change in percent correct, was larger for longer delays, up to 5% improvement, than for shorter delays. However, the variability of the measure is a function of the percent correct, and if the percent correct is closer to 100, the variance is lower. To enable a fair comparison of the performance change caused by stimulation as a function of delay length, we converted these percent change differences to changes in signal size using the signal detection theory metric d' (see STAR Methods), which measured how far above the noise the signal size, or memory, rose. This metric inferred the signal size from the behavioral responses, and it was normalized by the noise standard deviation. Using either signal detectability metrics or percent correct, no statistically significant changes in stimulation effect were found as a function of delay length using ANOVAs ($p > 0.1$). Stimulation improved performance across the entire working memory delay curve, and the change in d' ranged from 0.13 to 0.39, with an average at the longest delay of 0.16. Alternately, if the animals were at zero delay and 96% correct, improvements were 1.4%–1.7%, while at longer delays animals improved 3.6%–4% above 75%. Changes of this magnitude reached statistical significance in two to four behavioral sessions.

Insight into the speed with which these effects occurred after stimulation began was determined by analyzing the performance shortly after a condition change. If effects appeared slowly, then performance shortly after stimulation began should be lower than performance later in the trial block. Similarly, if the effects decreased slowly once stimulation stopped, the no-stimulation blocks should have had lower performance later in the block. Performance compared to trial position in the block is shown in Figures S3A and S3B for continuous and intermittent stimulation. No significant trends based on position within the blocks were found, which indicated that effects occurred in less than 3 min, which was the average time to complete ten trials.

The Interactions of Stimulation Effects with Cholinergic Pharmacology

To test the hypothesis that the improvement of working memory performance after intermittent stimulation depended

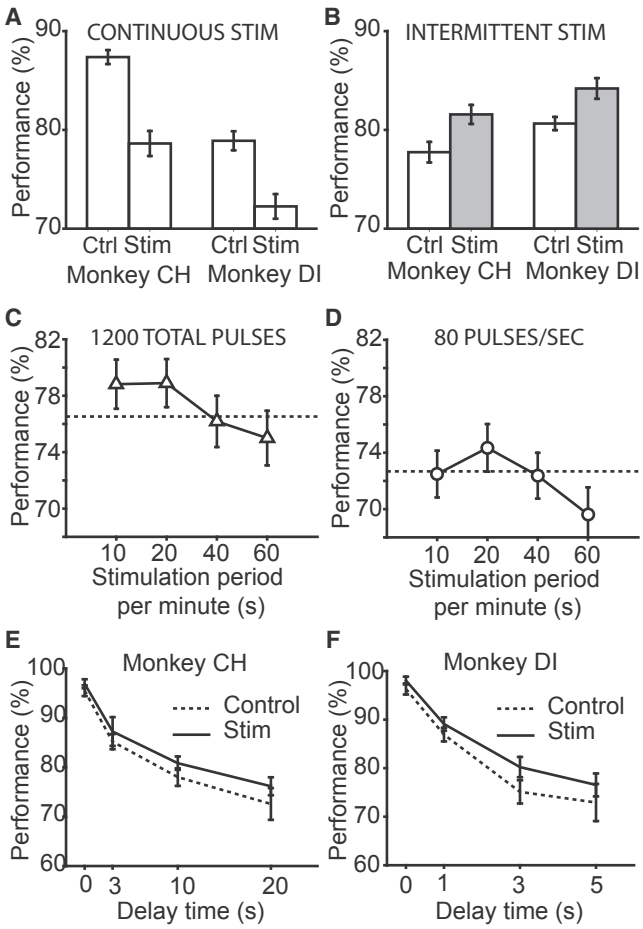


Figure 2. Effects of NB Stimulation on Working Memory Behavior

(A) The impact of continuous stimulation on performance. Monkeys performed the task with concurrent 80 Hz continuous stimulation or under a control condition.

(B) The impact of intermittent stimulation on performance. The gray bars indicate an intermittent stimulation condition distinct from the continuous stimulation condition indicated by empty bars in (A).

(C) Both monkeys were tested with 1,200 stimulation pulses per minute, delivered in 10 s (120 Hz), 20 s (60 Hz), 40 s (30 Hz), or 60 s (20 Hz). The performance under the control (no stimulation) condition for this week is indicated with the dashed line, and its standard error is 1.8%.

(D) Performance as both monkeys were tested with 80 pulses per second for different fractions of a minute. The standard error in control condition is 1.7%.

(E) Delay performance curve with and without intermittent stimulation in animal CH.

(F) Delay performance curve with and without intermittent stimulation in animal DI. Up to 1,000 trials were used to determine each point.

Error bars indicate the standard error of the mean. See also Figures S3 and S4.

on cholinergic pathways, stimulation was delivered with and without the cholinesterase inhibitor Donepezil. When the Donepezil dose was combined with intermittent stimulation, performance was no different from Donepezil alone, or stimulation alone, as shown in Figure 3A. The impact of combining Donepezil with 80 pulses per second continuous stimulation is shown on the right in Figure 3A. Continuous stimulation significantly impaired performance, but Donepezil reversed this impairment and resulted in supranormal performance. In Figure 3A, all comparisons relative to the control condition were significant (binomial test, $n > 3{,}000$ for each condition, $p < 0.005$ each). The result suggested that intermittent stimulation achieved a ceiling effect on cholinergic pathways. It also suggested that continuous stimulation degraded behavior by decreasing acetylcholine levels.

A further test of the cholinergic basis of effects was conducted using acetylcholine receptor antagonists. Mecamylamine was used to block nicotinic acetylcholine receptors in one set of experiments, and scopolamine was used to block muscarinic acetylcholine receptors in the second set of experiments. In either case (Figure 3B), intermittent stimulation did not improve performance ($p > 0.1$). The result suggested that both receptor subtypes were necessary for the performance increase caused by intermittent stimulation. As shown in Figure 3C, administration of mecamylamine and scopolamine alone reproduced known effects [16, 17]. Scopolamine impaired working memory (binomial test, $n = 550$, $p < 0.001$). Mecamylamine, at the dose tested, improved working memory (binomial test, $n = 2{,}700$, $p < 0.001$).

Donepezil is reported to cause side effects including nausea, which limits its clinical efficacy in some patients [8, 18, 19]. We plotted food consumption of the animals under control conditions, days in which intermittent stimulation was delivered in blocks, and days in which Donepezil was given. The results (Figure 3D) showed a significant effect of condition (unbalanced ANOVA; $F(2,56) = 3.4$ for condition, $p = 0.04$; $F(1,56) = 1.85$ for animal, $p = 0.179$). Post hoc testing found that monkeys consumed less food on days in which Donepezil was given relative to days in which stimulation was given (t test, $t = 2.81$, $df = 35$, $p < 0.009$).

The reaction times of animals CH and DI to the cue and match are shown in Figure S4. These reaction times were measured in the same data shown in Figures 2E and 2F. Under no conditions were there significant effects of stimulation on reaction times (paired t tests, $p > 0.1$ for all), which argued against stimulation causing a general effect on arousal or alertness.

Prior work stimulating NB in anesthetized animals has demonstrated consistent spectrum shifts of cortical electroencephalogram (EEG) recordings [20–23]. The low-frequency portions of the EEG, under 12 Hz, tend to decrease in energy, or desynchronize, while the higher-frequency gamma range increases under some experimental conditions. We recorded the subcortical local field potential (LFP), via the stimulating electrode, immediately before and after stimulation during periods of intermittent and continuous stimulation in the awake animal. As shown in Figures S2A–S2D, consistent reductions in LFP power spectrum occurred as a result of intermittent stimulation. Similar effects occurred after continuous stimulation at 80 Hz, shown in Figures S2E and S2F. All effects were significant (t tests, $p < 0.001$, frequency bands <4 Hz, 4–8 Hz, 8–12 Hz, 12–20 Hz, 20–40 Hz).

Long-Term Changes Caused by Stimulation

Unexpected improvements in working memory performance were observed in the three animals that were tested with stimulation of NB for months. In each animal, the delay interval was adjusted to maintain performance between 75% and 80%. The first indication of this change is shown in Figure 4A, in animal CH. The working memory delay duration had been increased

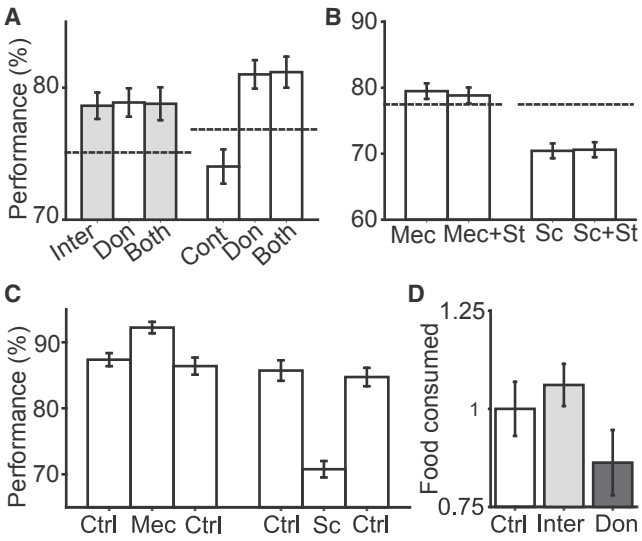


Figure 3. Pharmacological Interactions with Stimulation, and Effects on the Animals' Appetites

(A) On the left, the experimental conditions of intermittent stimulation (Inter) are compared to control (dashed line), Donepezil (Don), and both stimulation and Donepezil (Both). On the right, in a separate experiment, the same comparisons are applied with continuous stimulation (Cont). The gray bars indicate intermittent stimulation involved. Standard errors for control conditions (dashed lines) on the left and right are 1% and 1.1%, respectively.

(B) Effects of stimulation and acetylcholine receptor blockers. Mecamylamine (Mec) was given to animals prior to behavior, and working memory performance was assessed during intermittent stimulation and without stimulation. Scopolamine (Sc) was similarly tested. Average of the two animals tested in each condition is displayed. Dashed line shows the baseline performance in control condition, and standard error is 1.6%.

(C) Effects of Mecamylamine and Scopolamine. Animals performed behavior in control condition (left Ctrl), then were given Mecamylamine (Mec), or Scopolamine (Sc). Then animals performed washout behavior (right Ctrl, 72 hr after drug application).

(D) Effects of stimulation and Donepezil on appetite. Individual animal consumption under all conditions was normalized by mean control consumption to allow pooling across animals for this plot. Food consumption data were collected from animals CH and DI.

Error bars indicate the standard error of the mean.

from 6 s to 30 s over 9 weeks. Animal CH was improving shortly before stimulation was introduced, so further improvement could have been caused by a long-term stimulation effect or a continuation of an improvement trend.

Figures 2E and 2F give a chance to convert stimulation effects into equivalent changes in delay length. If the working memory delay chosen caused an animal to perform close to 75% correct, and the delay were halved, performance would have increased 5%–6%. This increase is slightly larger than the impact of intermittent stimulation shown in Figure 2B. A 5-fold decrease in the length of the delay period corresponds to a 12%–14% increase in percent correct. In signal-detection terms, a 14% increase in performance, from 75% to 89%, corresponds to a d' change of 0.78, which is 4.9 times greater than the average acute d' change caused by stimulation at long delays in Figures 2E and 2F.

Animal DI was allowed to reach asymptotic performance in the absence of stimulation for a longer period, and performance was stable over 11 weeks with a working memory delay of 2–3 s. Over the ensuing 5 months, the working memory delay increased from 3 s to 10 s, which is shown in Figure 4B. Under the null hypothesis that working memory ability was stable, this change was highly significant (binomial model, see STAR Methods, $p < 10^{-5}$). Animal PU (Figure 4C) performed the task for 17 weeks prior to initiating stimulation. Its performance was stable in this pre-stimulation period. In the following 7 weeks, the 1.5 s delay increased to 8 s, more than 5-fold, which again was significant ($p < 10^{-5}$). The results from animals DI and PU showed that stimulation increased working memory duration. All of the performance data used in this analysis were collected under control conditions, i.e., without the acute boost in performance shown in Figure 2B. For all data shown in Figure 4, the stimulation delivered in the several-month-long period of stimulation was mixed while the experiments of Figures 2, 3, and 4 were conducted.

DISCUSSION

The delayed match-to-sample task was used because it is a well-studied test of executive function in the nonhuman primate [24]. It embodies several mnemonic processes that are critical to human cognition, including attention, stimulus discrimination, encoding, and memory for recent events [25]. The task is sensitive to cholinergic modulation, as indicated by its enhancement by cholinesterase inhibitors [26, 27], and to cholinergic receptor agonists [28, 29], and it is sensitive to impairments of performance by cholinergic antagonists [17, 30]. Intermittent stimulation of NB led to short-term improvement in the task, whereas continuous stimulation suppressed performance. Pharmacological administration of cholinesterase inhibitors and acetylcholine receptor blockers demonstrated that the effects occur through modulation of acetylcholine levels and require activation of both classes of acetylcholine receptors. Unlike cholinesterase inhibitors, appetite is not suppressed by cholinergic modulation through deep brain stimulation. Even though stimulation was delivered in varied amounts each week, each of three animals increased their working memory delay duration 3- to 5-fold in the months after stimulation began. Two of those animals had extended periods of stable task performance before stimulation began. Future work may address the specificity of these effects, because working memory only comprises one component of executive function.

A major unexpected finding in this study is the form of stimulation that produced optimal results: 60 pulses per second delivered for 20 s and followed by 40 s without stimulation. Deep brain stimulation's mechanisms have been a subject of debate [31]. Locally, a depolarization block and possibly activation of local inhibitory circuits dominates to prevent somatic activation, whereas recording and imaging evidence suggests that axonal efferent outflow is activated by each stimulation pulse. Accordingly, the NB axons should be activated each time a pulse is delivered. Why the intermittent pattern results in more cortical acetylcholine, as inferred from pharmacological experiments, than continuous stimulation is an obvious question for subsequent studies. We speculate that intermittent, not continuous, neuronal stimulation results in higher levels of acetylcholine synthesis and release. The rate-limiting step in acetylcholine synthesis is the uptake of choline from the synaptic cleft via the high-affinity choline transporter hCHT [32]. Before release,

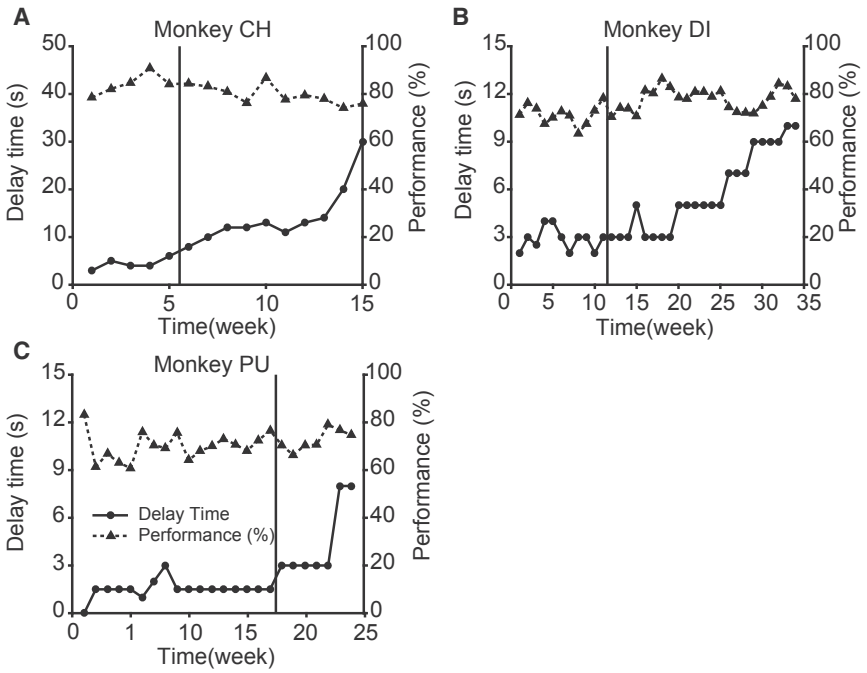


Figure 4. Changes in Working Memory over Time

(A) The working memory delay and performance for animal CH. Delay and performance are plotted against time relative to initiating intermittent stimulation. The delay was chosen to maintain performance range. The vertical line before week 6 indicates when intermittent stimulation was introduced. Left axis illustrates the delay, which is marked by circle symbols, while the right axis illustrates the performance, plotted as triangles. All data in this figure are defined under no-stimulation control conditions.

(B) Data for animal DI.

(C) Data for animal PU.

the proton-exchanging acetylcholine transporter VAChT brings more acetylcholine into the vesicle, and the high-affinity choline transporter, which would transport choline to the cytosol, is turned off by the low vesicular pH [33, 34]. After release, the vesicular membrane becomes part of the cell membrane, and the decreased proton concentrations turn the acetylcholine transporter off and the choline transporter on. If frequent vesicular release from high-frequency stimulation alters post-release activation of the high-affinity choline transporter by changing local pH, overall acetylcholine synthesis and release could be reduced.

A similar curiosity is that the stimulation does not have any temporal relation to the behavioral task or to known cholinergic kinetics other than the recycle time [35, 36]. Donepezil's effects on behavioral performance are arguably boosting the efficacy after normal cholinergic release, which would preserve the normal temporal function of acetylcholine [26]. However, some experiments with subtype-selective agonists also improve working memory function [37–39], and these would similarly lack any temporal specificity.

Improvement in cognition from cholinergic system activation could be caused by cholinergic modulation of neurons or by cholinergic-induced increases in blood flow, and these mechanisms are not mutually exclusive. In the neocortex, cholinergic activation of neurons leads to increased thalamic input into layer IV while suppressing other cortical inputs [40, 41]. This, at least partly, explains how the activation of the cholinergic system desynchronizes the EEG and local field potential (Figure S2), which was also reported by previous studies [21, 22, 42, 43]. The coupling of desynchronization to behavioral improvements suggests that direct cholinergic modulation of neurons improves the behavior.

At the same time, we cannot exclude the possibility that indirect mechanisms such as increases of blood flow played a role in the improved cognitive performance. In the brain, acetylcholine is acting on both neurons and glia adjacent to blood vessels [44, 45]. Cholinesterase inhibitors in humans cause an increase of 10%–15% in cerebral blood flow [46], and this blood flow is typically reduced in Alzheimer's patients [47–49]. Reductions in cerebral blood flow likely contribute to the cognitive impairments [50]. Future work may shed light on the relative contributions of neural modulation and blood flow regulation in these effects, as well as on the question of their independence.

The long-term effects observed indicate that working memory duration is increased 3- to 5-fold in each of three animals tested over periods of up to 5 months. The changes even occurred after long stable performance periods prior to stimulation. These results suggest that cholinergic modulation combined with executive function behavior causes brain plasticity to support behavioral improvements. Prior work has associated phasic stimulation of NB with neural plasticity in the sensory cortex [20, 22, 51].

Several clinical trials have been conducted on deep brain stimulation of cholinergic pathways [52, 53]. None have found consistent cognitive improvement using continuous stimulation. Similarly, animal work with continuous stimulation finds weak to no effects on learning and memory [11]. We measured analogous effects using continuous stimulation, which also suggest that high-frequency continuous stimulation suppresses acetylcholine levels. Our data show that making the stimulation intermittent is critical to turning cognitive suppression into cognitive enhancement. Animals had stronger appetites while receiving stimulation than while receiving Donepezil. The use of the intermittent stimulation parameters outlined in this work reach efficacies as high as those of high doses of cholinesterase inhibitors without peripheral side effects and point toward new candidates for therapeutic treatment of Alzheimer's dementia.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING

- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animal Model
- METHOD DETAILS
  - Behavioral Task
  - Surgery and Deep Brain Stimulation
  - Pharmacology and Stimulation Studies Administration
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - d' Calculation
  - Binomial Statistical Model
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2017.07.021.

AUTHOR CONTRIBUTIONS

Conceptualization and Funding Acquisition: D.T.B and C.C.; Project Administration: D.T.B.; Pharmacology Methodology: A.V.T. and P.M.C.; Investigation and Analysis: R.L., J.C., and D.T.B. performed experiments; Software: R.L. and D.T.B.; Writing: D.T.B., C.C., and R.L.; Review and Editing: all authors.

ACKNOWLEDGMENTS

Veterinary support provided by N. Rodriguez and P. Otovic. Artwork in Figure 1 was done by D. Bliss. This work funded by NIH grant 5R01MH097695.

Received: April 18, 2017
Revised: June 7, 2017
Accepted: July 11, 2017
Published: August 17, 2017

REFERENCES

1. Bartus, R.T. (2000). On neurodegenerative diseases, models, and treatment strategies: lessons learned and lessons forgotten a generation following the cholinergic hypothesis. Exp. Neurol. *163*, 495–529.
2. Terry, A.V., Jr., and Buccafusco, J.J. (2003). The cholinergic hypothesis of age and Alzheimer's disease-related cognitive deficits: recent challenges and their implications for novel drug development. J. Pharmacol. Exp. Ther. *306*, 821–827.
3. Sarter, M., and Bruno, J.P. (1997). Cognitive functions of cortical acetylcholine: toward a unifying hypothesis. Brain Res. Brain Res. Rev. *23*, 28–46.
4. Birks, J.S. (2006). Cholinesterase inhibitors for Alzheimer's disease. In Cochrane Database of Systematic Reviews, J.S. Birks, ed. (Chichester, UK: John Wiley & Sons, Ltd).
5. Hogan, D.B., Bailey, P., Carswell, A., Clarke, B., Cohen, C., Forbes, D., Man-Son-Hing, M., Lanctôt, K., Morgan, D., and Thorpe, L. (2007). Management of mild to moderate Alzheimer's disease and dementia. Alzheimers Dement. *3*, 355–384.
6. Qaseem, A., Snow, V., Cross, J.T., Jr., Forciea, M.A., Hopkins, R., Jr., Shekelle, P., Adelman, A., Mehr, D., Schellhase, K., Campos-Outcalt, D., et al.; American College of Physicians/American Academy of Family Physicians Panel on Dementia (2008). Current pharmacologic treatment of dementia: a clinical practice guideline from the American College of Physicians and the American Academy of Family Physicians. Ann. Intern. Med. *148*, 370–378.
7. Sabbagh, M., and Cummings, J. (2011). Progressive cholinergic decline in Alzheimer's Disease: consideration for treatment with donepezil 23 mg in patients with moderate to severe symptomatology. BMC Neurol. *11*, 21.
8. Farlow, M.R., Salloway, S., Tariot, P.N., Yardley, J., Moline, M.L., Wang, Q., Brand-Schieber, E., Zou, H., Hsu, T., and Satlin, A. (2010). Effectiveness and tolerability of high-dose (23 mg/d) versus standard-dose (10 mg/d) donepezil in moderate to severe Alzheimer's disease: A 24-week, randomized, double-blind study. Clin. Ther. *32*, 1234–1251.
9. Hendry, S.H.C., Jones, E.G., Killackey, H.P., and Chalupa, L.M. (1987). Choline acetyltransferase-immunoreactive neurons in fetal monkey cerebral cortex. Brain Res. *465*, 313–317.
10. Mesulam, M.M., Mufson, E.J., Levey, A.I., and Wainer, B.H. (1983). Cholinergic innervation of cortex by the basal forebrain: cytochemistry and cortical connections of the septal area, diagonal band nuclei, nucleus basalis (substantia innominata), and hypothalamus in the rhesus monkey. J. Comp. Neurol. *214*, 170–197.
11. Lee, J.E., Jeong, D.U., Lee, J., Chang, W.S., and Chang, J.W. (2016). The effect of nucleus basalis magnocellularis deep brain stimulation on memory function in a rat model of dementia. BMC Neurol. *16*, 6.
12. Pinto, L., Goard, M.J., Estandian, D., Xu, M., Kwan, A.C., Lee, S.-H., Harrison, T.C., Feng, G., and Dan, Y. (2013). Fast modulation of visual perception by basal forebrain cholinergic neurons. Nat. Neurosci. *16*, 1857–1863.
13. Selden, N.R., Gitelman, D.R., Salamon-Murayama, N., Parrish, T.B., and Mesulam, M.-M. (1998). Trajectories of cholinergic pathways within the cerebral hemispheres of the human brain. Brain *121*, 2249–2257.
14. Rohlfing, T., Kroenke, C.D., Sullivan, E.V., Dubach, M.F., Bowden, D.M., Grant, K.A., and Pfefferbaum, A. (2012). The INIA19 template and NeuroMaps atlas for primate brain image parcellation and spatial normalization. Front. Neuroinform. *6*, 27.
15. McIntyre, C.C., Grill, W.M., Sherman, D.L., and Thakor, N.V. (2004). Cellular effects of deep brain stimulation: model-based analysis of activation and inhibition. J. Neurophysiol. *91*, 1457–1469.
16. Terry, A.V., Buccafusco, J.J., and Prendergast, M.A. (1999). Dose-specific improvements in memory-related task performance by rats and aged monkeys administered the nicotinic-cholinergic antagonist mecamylamine. Drug Dev. Res. *47*, 127–136.
17. Buccafusco, J.J., Terry, A.V., Jr., Webster, S.J., Martin, D., Hohnadel, E.J., Bouchard, K.A., and Warner, S.E. (2008). The scopolamine-reversal paradigm in rats and monkeys: the importance of computer-assisted operant-conditioning memory tasks for screening drug candidates. Psychopharmacology (Berl.) *199*, 481–494.
18. Yoshida, T., Ha-Kawa, S., Yoshimura, M., Nobuhara, K., Kinoshita, T., and Sawada, S. (2007). Effectiveness of treatment with donepezil hydrochloride and changes in regional cerebral blood flow in patients with Alzheimer's disease. Ann. Nucl. Med. *21*, 257–265.
19. Han, S.-H., Lee, J.-H., Kim, S.Y., Park, K.W., Chen, C., Tripathi, M., Dash, A., and Kubota, N. (2017). Donepezil 23 mg in Asian patients with moderate-to-severe Alzheimer's disease. Acta Neurol. Scand. *135*, 252–256.
20. Bakin, J.S., and Weinberger, N.M. (1996). Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. Proc. Natl. Acad. Sci. USA *93*, 11219–11224.
21. Bjordahl, T.S., Dimyan, M.A., and Weinberger, N.M. (1998). Induction of long-term receptive field plasticity in the auditory cortex of the waking guinea pig by stimulation of the nucleus basalis. Behav. Neurosci. *112*, 467–479.
22. Kilgard, M.P., and Merzenich, M.M. (1998). Cortical map reorganization enabled by nucleus basalis activity. Science *279*, 1714–1718.
23. McLin, D.E., 3rd, Miasnikov, A.A., and Weinberger, N.M. (2002). Induction of behavioral associative memory by stimulation of the nucleus basalis. Proc. Natl. Acad. Sci. USA *99*, 4002–4007.
24. Riley, M.R., and Constantinidis, C. (2016). Role of prefrontal persistent activity in working memory. Front. Syst. Neurosci. *9*, 181.
25. Rodriguez, J.S., and Paule, M.G. (2009). Working memory delayed response tasks in monkeys. In Methods Behav. Anal. Neurosci., Second Edition, J.J. Buccafusco, ed. (Boca Raton: CRC Press/Taylor & Francis).
26. Buccafusco, J.J., and Terry, A.V. (2004). Donepezil-induced improvement in delayed matching accuracy by young and old rhesus monkeys. J. Mol. Neurosci. *24*, 85–91.

27. Callahan, P.M., Hutchings, E.J., Kille, N.J., Chapman, J.M., and Terry, A.V., Jr. (2013). Positive allosteric modulator of α7 nicotinic-acetylcholine receptors, PNU-120596 augments the effects of donepezil on learning and memory in aged rodents and non-human primates. Neuropharmacology *67*, 201–212.
28. Terry, A.V., Jr., Plagenhoef, M., and Callahan, P.M. (2016). Effects of the nicotinic agonist varenicline on the performance of tasks of cognition in aged and middle-aged rhesus and pigtail monkeys. Psychopharmacology (Berl.) *233*, 761–771.
29. Terry, A.V., Jr., Risbrough, V.B., Buccafusco, J.J., and Menzaghi, F. (2002). Effects of (+/-)-4-[[2-(1-methyl-2-pyrrolidinyl)ethyl]thio]phenol hydrochloride (SIB-1553A), a selective ligand for nicotinic acetylcholine receptors, in tests of visual attention and distractibility in rats and monkeys. J. Pharmacol. Exp. Ther. *301*, 284–292.
30. Zhou, X., Qi, X.-L., Douglas, K., Palaninathan, K., Kang, H.S., Buccafusco, J.J., Blake, D.T., and Constantinidis, C. (2011). Cholinergic modulation of working memory activity in primate prefrontal cortex. J. Neurophysiol. *106*, 2180–2188.
31. Vitek, J.L. (2002). Mechanisms of deep brain stimulation: excitation or inhibition. Mov. Disord. *17* (*Suppl 3*), S69–S72.
32. Okuda, T., Haga, T., Kanai, Y., Endou, H., Ishihara, T., and Katsura, I. (2000). Identification and characterization of the high-affinity choline transporter. Nat. Neurosci. *3*, 120–125.
33. Iwamoto, H., Blakely, R.D., and De Felice, L.J. (2006). Na+, Cl-, and pH dependence of the human choline transporter (hCHT) in Xenopus oocytes: the proton inactivation hypothesis of hCHT in synaptic vesicles. J. Neurosci. *26*, 9851–9859.
34. Nguyen, M.L., Cox, G.D., and Parsons, S.M. (1998). Kinetic parameters for the vesicular acetylcholine transporter: two protons are exchanged for one acetylcholine. Biochemistry *37*, 13400–13410.
35. Trabucchi, M., Cheney, D.L., Hanin, I., and Costa, E. (1975). Application of principles of steady-state kinetics to the estimation of brain acetylcholine turnover rate: effects of oxotremorine and physostigmine. J. Pharmacol. Exp. Ther. *194*, 57–64.
36. Cheney, D.L., Trabucchi, M., Racagni, G., Wang, C., and Costa, E. (1974). Effects of acute and chronic morphine on regional rat brain acetylcholine turnover rate. Life Sci. *15*, 1977–1990.
37. Terry, A.V., Jr., Buccafusco, J.J., Borsini, F., and Leusch, A. (2002). Memory-related task performance by aged rhesus monkeys administered the muscarinic M(1)-preferring agonist, talsaclidine. Psychopharmacology (Berl.) *162*, 292–300.
38. Castner, S.A., Smagin, G.N., Piser, T.M., Wang, Y., Smith, J.S., Christian, E.P., Mrzljak, L., and Williams, G.V. (2011). Immediate and sustained improvements in working memory after selective stimulation of α7 nicotinic acetylcholine receptors. Biol. Psychiatry *69*, 12–18.
39. Buccafusco, J.J., and Jackson, W.J. (1991). Beneficial effects of nicotine administered prior to a delayed matching-to-sample task in young and aged monkeys. Neurobiol. Aging *12*, 233–238.
40. Metherate, R., and Ashe, J.H. (1993). Nucleus basalis stimulation facilitates thalamocortical synaptic transmission in the rat auditory cortex. Synapse *14*, 132–143.
41. Aramakis, V.B., and Metherate, R. (1998). Nicotine selectively enhances NMDA receptor-mediated synaptic transmission during postnatal development in sensory neocortex. J. Neurosci. *18*, 8485–8495.
42. Bakin, J.S., South, D.A., and Weinberger, N.M. (1996). Induction of receptive field plasticity in the auditory cortex of the guinea pig during instrumental avoidance conditioning. Behav. Neurosci. *110*, 905–913.
43. McLin, D.E., 3rd, Miasnikov, A.A., and Weinberger, N.M. (2002). The effects of electrical stimulation of the nucleus basalis on the electroencephalogram, heart rate, and respiration. Behav. Neurosci. *116*, 795–806.
44. Vaucher, E., Borredon, J., Seylaz, J., and Lacombe, P. (1995). Autoradiographic distribution of cerebral blood flow increases elicited by stimulation of the nucleus basalis magnocellularis in the unanesthetized rat. Brain Res. *691*, 57–68.
45. Vaucher, E., Linville, D., and Hamel, E. (1997). Cholinergic basal forebrain projections to nitric oxide synthase-containing neurons in the rat cerebral cortex. Neuroscience *79*, 827–836.
46. Lojkowska, W., Ryglewicz, D., Jedrzejczak, T., Minc, S., Jakubowska, T., Jarosz, H., and Bochynska, A. (2003). The effect of cholinesterase inhibitors on the regional blood flow in patients with Alzheimer's disease and vascular dementia. J. Neurol. Sci. *216*, 119–126.
47. Sharp, P., Gemmell, H., Cherryman, G., Besson, J., Crawford, J., and Smith, F. (1986). Application of iodine-123-labeled isopropylamphetamine imaging to the study of dementia. J. Nucl. Med. *27*, 761–768.
48. Jagust, W.J., Budinger, T.F., and Reed, B.R. (1987). The diagnosis of dementia with single photon emission computed tomography. Arch. Neurol. *44*, 258–262.
49. Schuff, N., Matsumoto, S., Kmiecik, J., Studholme, C., Du, A., Ezekiel, F., Miller, B.L., Kramer, J.H., Jagust, W.J., Chui, H.C., and Weiner, M.W. (2009). Cerebral blood flow in ischemic vascular dementia and Alzheimer's disease, measured by arterial spin-labeling magnetic resonance imaging. Alzheimers Dement. *5*, 454–462.
50. Ruitenberg, A., den Heijer, T., Bakker, S.L.M., van Swieten, J.C., Koudstaal, P.J., Hofman, A., and Breteler, M.M.B. (2005). Cerebral hypoperfusion and clinical onset of dementia: the Rotterdam Study. Ann. Neurol. *57*, 789–794.
51. Bao, S., Chang, E.F., Davis, J.D., Gobeske, K.T., and Merzenich, M.M. (2003). Progressive degradation and subsequent refinement of acoustic representations in the adult auditory cortex. J. Neurosci. *23*, 10765–10775.
52. Laxton, A.W., Tang-Wai, D.F., McAndrews, M.P., Zumsteg, D., Wennberg, R., Keren, R., Wherrett, J., Naglie, G., Hamani, C., Smith, G.S., and Lozano, A.M. (2010). A phase I trial of deep brain stimulation of memory circuits in Alzheimer's disease. Ann. Neurol. *68*, 521–534.
53. Kuhn, J., Hardenacke, K., Lenartz, D., Gruendler, T., Ullsperger, M., Bartsch, C., Mai, J.K., Zilles, K., Bauer, A., Matusch, A., et al. (2015). Deep brain stimulation of the nucleus basalis of Meynert in Alzheimer's dementia. Mol. Psychiatry *20*, 353–360.
54. Mesulam, M.-M., Mufson, E.J., Levey, A.I., and Wainer, B.H. (1984). Atlas of cholinergic neurons in the forebrain and upper brainstem of the macaque based on monoclonal choline acetyltransferase immunohistochemistry and acetylcholinesterase histochemistry. Neuroscience *12*, 669–686.
55. Gielow, M.R., and Zaborszky, L. (2017). The input-output relationship of the cholinergic basal forebrain. Cell Rep. *18*, 1817–1830.
56. Paxinos, G., and Huang, X.F.A.W.T. (2000). The Rhesus Monkey Brain in Stereotaxic Coordinates (San Diego: Academic Press).
57. McCairn, K.W., and Turner, R.S. (2015). Pallidal stimulation suppresses pathological dysrhythmia in the parkinsonian motor cortex. J. Neurophysiol. *113*, 2537–2548.
58. McCairn, K.W., and Turner, R.S. (2009). Deep brain stimulation of the globus pallidus internus in the parkinsonian primate: local entrainment and suppression of low-frequency oscillations. J. Neurophysiol. *101*, 1941–1960.
59. Tsukada, H., Nishiyama, S., Fukumoto, D., Ohba, H., Sato, K., and Kakiuchi, T. (2004). Effects of acute acetylcholinesterase inhibition on the cerebral cholinergic neuronal system and cognitive function: Functional imaging of the conscious monkey brain using animal PET in combination with microdialysis. Synapse *52*, 1–10.
60. Green, D.M., and Swets, J.A. (1966). Signal Detection Theory and Psychophysics (New York: Wiley).
61. Smith, D.G., and Duncan, M.J.J. (2004). Testing theories of recognition memory by predicting performance across paradigms. J. Exp. Psychol. Learn. Mem. Cogn. *30*, 615–625.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Donepezil Hydrochloride	Memory Pharmaceuticals, Montvale, NJ	N/A
Scopolamine Hydrobromide	Sigma Aldrich	CAS # 114-49-8
Mecamylamine Hydrochloride	Sigma Aldrich	CAS # 826-39-1
Deposited Data		
http://dx.doi.org/10.6084/m9.figshare.5039153	N/A	N/A
Experimental Models: Organisms/Strains		
Macaca mulatta	Mannheimer Foundation	N/A
Software and Algorithms		
MATLAB R2016a	MathWorks	N/A
DTS_Gracewood	PBTLI	http://www.pbtli.com
d' calculation	Smith and Duncan 2004	N/A
Binomial model	This paper	http://dx.doi.org/10.6084/m9.figshare.5039153

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to, and will be fulfilled by, the Lead Contact, David T Blake (dblake.mcg@gmail.com).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal Model

Five rhesus monkeys (*Macaca mulatta*) were used in this study. Two initial animals (K and S) were used to optimize the placing of the stimulation electrodes. These were a 7 years old male and a 12 years old female weighing 7 and 6 kg. The three subsequent animals (animals CH, DI, and PU), used for data in Figures 2–4, were male, 6 years old and weighed 8 −11 kg. The monkeys were pair-housed in 12-h light/dark cycle rooms. All monkeys were task naive at the start of the experiments. All animals were chaired during task performance, with one arm restrained to both establish consistency in the arm used to perform the task, and to prevent spinning in the primate chair. Animals were not head fixed during task performance.

All animal studies comply with the Guide for the Care and Use of Animals, 8th Edition, and were approved by the IACUC at Augusta University.

METHOD DETAILS

Behavioral Task

All behavioral software was based on prior work [28], and available from PBTLI Prime Behavior Testing Laboratories, Inc, http://www.pbtli.com/, Augusta, GA. The behavioral task, shown in Figure 1, initiated with a colored square cue in the center-top part of the touchscreen. After the animal touched the square, the cue disappeared, and the screen remained blank for the delay period. In the match phase, two colored squares were presented in the lower right, and lower left, of the touchscreen. One of the two squares, randomized in location, had the same color as the cue square. Correctly touching the target square resulted in delivery of food slurry reward, while incorrect responses resulted in doubling the inter trial period as a brief timeout. The animal reaction times were not strongly constrained. Animals had 10 s to make a response. Percent correct and reaction times were recorded. Animals required several months of daily training to achieve stable performance at the task. Three colors (red, blue and yellow) were used. In each trial, two were randomly chosen as cue/match, and as the distractor.

Dates of data collection for each animal for each figure.

Experiment	Animal CH	Animal DI	Animal PU
Figures 2A, 2B, and S3	8/15-10/15	11/15-1/16	
Figures 2C and 2D	6/16-7/16	6/16-6/16	
Figures 2E, 2F, and S4	10/15-11/15	11/15-12/15	
Figure 3	2/16-8/16	6/16-8/16	8/2016
Figure 4	7/15-11/15	3/15-11/15	12/15-8/16

Surgery and Deep Brain Stimulation

Surgery was performed in a sterile surgical suite under isoflurane anesthesia monitored by a clinical veterinarian. Stereotaxic measurements were made to target implantation. In our first two animals, targeting was 8mm lateral, 12 and 14 mm anterior interaural respectively, and 29 mm below the cortical surface in a vertical penetration. The final three animals were implanted at 16 mm anterior interaural with the same lateral and depth coordinates. The final lateral and anterior coordinates, and depth, were chosen to correspond to the center of the anterior portion of the Nucleus Basalis of Meynert, which would contain the highest density of projections to the prefrontal cortex [10, 54, 55]. At the targeted brain position, the stimulation region, which is estimated to be a 4 mm diameter sphere [31], should include the Nucleus Basalis of Meynert as well as portions of the anterior amygdala including the central nucleus, the anterior commissure, and the inferior internal globus pallidus [56]. A skin flap was performed, and two holes were drilled above the location of each implant. A small cylindrical titanium chamber (5-mm inner diameter and 7-mm outer) was mounted on the cranium and the skin was sutured into place leaving the chamber exposed. A 26 ga. hypodermic guide tube was lowered and the tip advanced 5 mm below the dura mater. The electrode was inserted into the guide tube, and a stylus was used to push it to the appropriate depth. The use of the guide tube and stylus followed published methods [57, 58], which are analogous to methods used in human neurosurgical deep brain electrode implantation. The guide tube was then raised while the stylus depth maintained. The chamber was evacuated of fluid, flushed with ceftriaxone, and thereafter fluid evacuated a second time. Silicone was poured into the chamber to seal the fenestrations in the skull and the inside of the chamber. The electrode was fixed in depth with a drop of cyanoacrylate, and its rear wire was stripped and soldered to a connector that was fixed on the chamber outer wall. One week after the surgery, the animals returned to behavioral studies.

The stimulation pulses were created by a multiple functional I/O device (USB-6211, National Instruments, Austin, TX), which was controlled by a custom programed software. Impedances of electrodes were checked daily prior to each behavioral session so that stimulation voltages could be tailored to deliver the appropriate currents (200 μA). Continuous stimulation was performed from 20 pulses per second to 120. Intermittent stimulation was initially applied at 60 pulses per second only in-between trials, or only during trials, which were roughly 3 and 5 s periods. Thereafter, experiments standardized on intermittent stimulation using a one-minute cycle length, and stimulation for 20 s at 60 pulses per second, followed by 40 s with no stimulation. In these one-minute cycle stimulation experiments, no attempt was made to synchronize the stimulation with the behavioral trials.

Electrodes were custom manufactured in our laboratory based on published specifications [57, 58]. Conductors were 50 μm Pt/Ir, Teflon-insulated wire (A-M systems, Seattle, WA) embedded within a 30 ga. hypodermic tube, which was encased in a 28 ga. polyimide sheath. The wire extended from the end of the sheath into the brain tissue by roughly 1 mm, and the last 0.7 mm of insulation was stripped to achieve impedances of 5-10 kOhm at 1 kHz. The far end of the electrode was soldered to an extracranial connector fixed on the chamber outer wall. Preliminary experiments on electrode placement in the two pilot animals tested the effects of short periods of stimulation on EEG desynchronization. Stimulation was delivered with biphasic, negative first, unipolar 200 μA pulses with 100 μS per phase, and 10 pulses were delivered in 100 msec. Under anesthesia, this resulted in EEG desynchronization when averaged over 10 trials when the electrode was at a depth corresponding to the atlas position of Nucleus Basalis. EEG was recorded with × 1000 gain from two bone screws at the same M-L position, with one near vertex, and one lateralized. An electrode movement vertically in either direction of more than 1 mm was adequate to make desynchronization not possible using the same protocol. After our second pilot animal, we determined the depth for our A-P and M-L coordinates was 29 mm below the cortical surface. We further recovered our electrode track from our second pilot animal, shown in Figure S1. It was implanted 2 mm more caudally than the electrodes from the three monkeys presented in the results. The tip of the electrode was centered 1 mm medial to Nucleus Basalis at 14 mm anterior interaural. After our pilot animals, we implanted to the same depth in all animals, and 2 mm more rostrally, for the current study, and measured LFP desynchronization in the awake animals using the stimulation electrode as the positive lead (Figure S2).

In these first two pilot animals, we frequently moved electrodes vertically to new positions. At this point in the study, we understood that the electrode depth at which EEG desynchronization occurred after brief phasic stimulation precisely matched that at which 80 Hz continuous stimulation resulted in significant performance decrements. The rear ends of the electrodes were always visible, protruding from the implanted titanium cylinder, which gave an absolute depth measurement.

Pharmacology and Stimulation Studies Administration

We used donepezil, mecamylamine and scopolamine (Sigma-Aldrich, St. Louis, MO) in this study. The drugs were dissolved in a 0.5 cc saline vehicle to an amount specified by the experiments (donepezil: 200 μg/kg, mecamylamine: 300 μg/kg and scopolamine:

6.2 μg/kg). Donepezil was given via I.M. administration 15 min before behavior testing on the mornings in which data was collected. The rationale for the doses selected for Donepezil was based on previous behavioral and functional brain imaging data (dose range 50-250 μg/kg) in rhesus monkeys [27, 59]. Data from days without donepezil were at least 72 hr after the last administration. Doses higher than 200 μg/kg could not be used in this study, as the animals administered these doses refused to engage in behavioral trials for food reward. Nausea is a well-known, clinical side effect of the drug in humans. Each week, control performance was evaluated. Then, performances under the intermittent stimulation alone, donepezil alone, and donepezil+stimulation were evaluated on successive days. Similar weekly plans were used for continuous stimulation, and mecamylamine (300 μg/kg) and scopolamine (6.2 μg/kg) testing. This weekly testing schedule was chosen over an intraday comparison because an intraday design would compare drug versus no-drug conditions at different times, and with different appetitive motivations. The recovery test sessions from drug were recorded 72 hr after drug application. The doses of mecamylamine and scopolamine used in this study refer to previous publications [16, 30].

Initial deep brain stimulation studies, shown in Figures 2A and 2B, provided stimulation in blocks. For continuous stimulation, 100 trials without stimulation were interleaved with 100 trials with stimulation. For intermittent stimulation, blocks were 50 trials. For Figures 2C and 2D, all stimulation conditions were executed randomly block by block and interleaved with each other. Each block also contained 50 trials. For Figures 3A–3C, we collected animals' performance data in control conditions for 1-2 days first, then collected performance data after administering pharmacological agents. Drug and Stim+Drug conditions were interleaved with each other in blocks of 50 trials. The recording procedures for the pharmacology experiments were 2-3 times to improve statistical power. For Figures 2E, 2F, and S4, which includes performance at different delays with and without stimulation, we tested the animals' performance with interleaved blocks in control and stimulation conditions. In each block, variable delay times were pseudo-randomly chosen and arranged in the test sequence. Software was administered directly from PBTLI software while a second computer was used to generate stimulation pulses. Data for Figure 4 were selected from all available data in which no drug or stimulation condition occurred.

Delay times used in each experiment.

Experiment	Animal CH	Animal DI	Animal PU
Figures 2A & S3A	10 s	3 s	
Figures 2B & S3B	13 s	3 s	
Figures 2C and 2D	6 s	7 s	
Figures 2E, 2F, and S4	0 s, 3 s, 10 s, 20 s	0 s, 1 s, 3 s, 5 s	
Figure 3A left	6 s	5 s	
Figure 3A right	6 s	10 s	
Figure 3B	8 s		8 s
Figure 3C		10 s	3 s

QUANTIFICATION AND STATISTICAL ANALYSIS

d' Calculation

The signal detection theoretic d' is the salience of the target relative to the distractor as inferred from the animals' behavioral choices. In our two alternative forced choice framework, d' is calculated using only the percent correct with the MATLAB (Mathworks, Natick, MA) norminv function, as $\sqrt{2} * norminv(Hit\ Rate)$ [60, 61]. This metric measures the amount of the memory that is retained, on average. It assumes the memory can be measured in terms of a mean and a noise standard deviation. As time passes, that mean, or memory, regresses toward zero. The metric measures the magnitude, normalized by the noise standard deviation, of the mean at the end of the delay period. The following table details how d' relates to percent correct in the task.

d'	Percent Correct
0.25	57
0.5	64
0.75	70
1.0	76
1.25	81
1.5	85
1.75	89
2.0	92

Binomial Statistical Model

We relied on a binomial statistic model to test if Nucleus Basalis stimulation, over weeks and months, produced long-term improvements in working memory that could be measured which resulted in ability to perform the task at high performance with a longer delay interval (Figure 4). Prior to stimulation, animals performed 3000 trials weekly at a performance level averaging 78% correct, with a delay interval ranging between 1.5-6 s for the three animals. For each animal, we asked what performance improvement would occur if the delay, at which performance level is 78% correct, were doubled? In each animal's delay curves, shown in Figures 2E and 2F, halving the delay would result in changes in percent correct of 5.4 and 5.1 percent. In other words, if we time scaled the graphs in Figures 2E and 2F to the same percentage correct at twice the delays, animal performance at the original delay would be predicted, based on the delay curves in Figures 2E and 2F, to increase by just over 5%. We conservatively rounded this percentage to a 5 percent change. Next, the two-tailed probabilities were calculated that a binomial process with a fraction correct of 0.78 over 3000 trials could result in more correct trials than a process with a fraction correct 5 percent higher or lower. This probability is less than 10^{-5}. The implication of the binomial model is that if an animal performs at 78% correct for duration 2X, then it would be expected to perform at 83%, or higher, at duration X, and that this performance would be significantly better than its pre-stimulation performance of 78% at duration X.

DATA AND SOFTWARE AVAILABILITY

Data used to generate these figures are archived at http://dx.doi.org/10.6084/m9.figshare.5039153.

Making health habitual:
the psychology of 'habit-formation' and general practice

MAKING HEALTH HABITUAL

The Secretary of State recently proposed that the NHS:

'... take every opportunity to prevent poor health and promote healthy living by making the most of healthcare professionals' contact with individual patients.'[1]

Patients trust health professionals as a source of advice on 'lifestyle' (that is, behaviour) change, and brief opportunistic advice can be effective.[2] However, many health professionals shy away from giving advice on modifying behaviour because they find traditional behaviour change strategies time-consuming to explain and difficult for the patient to implement.[2] Furthermore, even when patients successfully initiate the recommended changes, the gains are often transient[3] because few of the traditional behaviour change strategies have built-in mechanisms for maintenance.

Brief advice is usually based on advising patients on what to change and why (for example, reducing saturated fat intake to reduce the risk of heart attack). Psychologically, such advice is designed to engage conscious deliberative motivational processes, which Kahneman terms 'slow' or 'System 2' processes.[4] However, the effects are typically short-lived because motivation and attention wane. Brief advice on how to change, engaging automatic ('System 1') processes, may offer a valuable alternative with potential for long-term impact.

Opportunistic health behaviour advice must be easy for health professionals to give and easy for patients to implement to fit into routine health care. We propose that simple advice on how to make healthy actions into habits — externally-triggered automatic responses to frequently encountered contexts — offers a useful option in the behaviour change toolkit. Advice for creating habits is easy for clinicians to deliver and easy for patients to implement: repeat a chosen behaviour in the same context, until it becomes automatic and effortless.

HABIT FORMATION AND HEALTH

While often used as a synonym for frequent or customary behaviour in everyday parlance, within psychology, 'habits' are defined as actions that are triggered automatically in response to contextual cues that have been associated with their performance:[5,6] for example, automatically washing hands (action) after using the toilet (contextual cue), or putting on a seatbelt (action) after getting into the car (contextual cue). Decades of psychological research consistently show that mere repetition of a simple action in a consistent context leads, through associative learning, to the action being activated upon subsequent exposure to those contextual cues (that is, habitually).[7–9] Once initiation of the action is 'transferred' to external cues, dependence on conscious attention or motivational processes is reduced.[10] Therefore habits are likely to persist even after conscious motivation or interest dissipates.[11] Habits are also cognitively efficient, because the automation of common actions frees mental resources for other tasks.

A growing literature demonstrates the relevance of habit-formation principles to health.[12,13] Participants in one study repeated a self-chosen health-promoting behaviour (for example, eat fruit, go for a walk) in response to a single, once-daily cue in their own environment (such as, after breakfast). Daily ratings of the subjective automaticity of the behaviour (that is, habit strength) showed an asymptotic increase, with an initial acceleration that slowed to a plateau after an average of 66 days.[9] Missing the occasional opportunity to perform the behaviour did not seriously impair the habit formation process: automaticity gains soon resumed after one missed performance.[9] Automaticity strength peaked more quickly for simple actions (for example, drinking water) than for more elaborate routines (for example, doing 50 sit-ups).

Habit-formation advice, paired with a 'small changes' approach, has been tested as a behaviour change strategy.[14,15] In one study, volunteers wanting to lose weight were randomised to a habit-based intervention, based on a brief leaflet listing 10 simple diet and activity behaviours and encouraging context-dependent repetition, or a no-treatment waiting list control. After 8 weeks, the intervention group had lost 2 kg compared with 0.4 kg in the control group. At 32 weeks, completers in the intervention group had lost an average of 3.8 kg.[14] Qualitative interview data indicated that automaticity had developed: behaviours became 'second nature', 'worming their way into your brain' so that participants 'felt quite strange' if they did not do them.[10] Actions that were initially difficult to stick to became easier to maintain. A randomised controlled trial is underway to test the efficacy of this intervention where delivered in a primary care setting to a larger sample, over a 24-month follow-up period.[16] Nonetheless, these early results indicate that habit-forming processes transfer to the everyday environment, and suggest that habit-formation advice offers an innovative technique for promoting long-term behaviour change.[13]

MAKING HEALTHY HABITS

We suggest that professionals could consider providing habit-formation advice as a way to promote long-term behaviour change among patients. Habit-formation advice is ultimately simple — repeat an action consistently in the same context.[12] The habit formation attempt begins at the 'initiation phase', during which the new behaviour and the context in which it will be done are selected. Automaticity develops in the subsequent 'learning phase', during which the behaviour is repeated in the chosen context to strengthen the context-behaviour association (here a simple ticksheet for self-monitoring performance may help; Box 1). Habit-formation culminates in the 'stability phase', at which the habit has formed and its strength has plateaued, so that it persists over time with minimal effort or deliberation.

"Advice for creating habits is easy for clinicians to deliver and easy for patients to implement: repeat a chosen behaviour in the same context, until it becomes automatic and effortless."

"... typical [behaviour change] *advice ... emphasises variation in behaviours and settings ... . Variation may stave off boredom, but is effortful and ... incompatible with development of automaticity."*

Initiation requires the patient to be sufficiently motivated to begin a habit-formation attempt, but many patients would like to eat healthier diets or take more exercise, for example, if doing so were easy. Patients must choose an appropriate context in which to perform the action. The 'context' can be any cue, for example, an event ('when I get to work') or a time of day ('after breakfast'), that is sufficiently salient in daily life that it is encountered and detected frequently and consistently. A cue located within an existing daily routine (for example, 'when I go on my lunch break') provides a convenient and stable starting point.[10]

Keeping going during the learning phase is crucial. The idea of repeating a single specific action (for example, eating a banana) in a consistent context (with cereal at breakfast) is very different from typical advice given to people trying to take up new healthy behaviours, which often emphasises variation in behaviours and settings to maintain interest (trying different fruits with or between different meals). Variation may stave off boredom, but is effortful and depends on maintaining motivation, and is incompatible with development of automaticity.[6]

Patients should choose the target behaviour themselves. Progress towards a self-determined behavioural goal supports patients' sense of autonomy and sustains interest,[17] and there is evidence that a behaviour change selected on the basis of its personal value, rather than to satisfy external demands such as physicians' recommendations, is an easier habit target.[18] Patients need to select a new behaviour (for example, eat an apple) rather than give up an existing behaviour (do not eat fried snacks) because it is not possible to form a habit for not doing something. The automaticity of habit means that breaking existing habits requires different and altogether more effortful strategies than making new habits.[12]

Patients should be encouraged to aim for small and manageable behaviour changes, because failure can be discouraging. A sedentary person, for example, would be more appropriately advised to walk one or two stops more before getting on the bus than to walk the entire route — at least for their first habit goal. Small changes can benefit health: slight adjustments to dietary intake can aid long-term weight management,[19] and small amounts of light physical activity are more beneficial than none.[20] Moreover, simpler actions become habitual more quickly.[9] Additionally, behaviour change achievements, however small, can increase self-efficacy, which can in turn stimulate pursuit of further changes.[21] Forming one 'small' healthy habit may thereby increase self-confidence for working towards other health-promoting habits.

Unrealistic expectations of the duration of the habit formation process can lead the patient to give up during the learning phase. Some patients may have heard that habits take 21 days to form. This myth appears to have originated from anecdotal evidence of patients who had received plastic surgery treatment and typically adjusted psychologically to their new appearance within 21 days.[22] More relevant research found that automaticity plateaued on average around 66 days after the first daily performance,[9] although there was considerable variation across participants and behaviours. Therefore, it may be helpful to tell patients to expect habit formation (based on daily repetition) to take around

Box 1. A tool for patients

Make a new healthy habit

1. Decide on a goal that you would like to achieve for your health.
2. Choose a simple action that will get you towards your goal which you can do on a daily basis.
3. Plan when and where you will do your chosen action. Be consistent: choose a time and place that you encounter every day of the week.
4. Every time you encounter that time and place, do the action.
5. It will get easier with time, and within 10 weeks you should find you are doing it automatically without even having to think about it.
6. Congratulations, you've made a healthy habit!

My goal (e.g. 'to eat more fruit and vegetables') _____

My plan (e.g. 'after I have lunch at home I will have a piece of fruit')
(When and where) _____ I will _____

Some people find it helpful to keep a record while they are forming a new habit. This daily tick-sheet can be used until your new habit becomes automatic. You can rate how automatic it feels at the end of each week, to watch it getting easier.

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10
Monday										
Tuesday										
Wednesday										
Thursday										
Friday										
Saturday										
Sunday										
Done on >5 days, yes or no										
How automatic does it feel? Rate from 1 (not at all) to10 (completely)										

ADDRESS FOR CORRESPONDENCE

Benjamin Gardner
Health Behaviour Research Centre, Department of Epidemiology and Public Health, University College London, Gower Street, London, WC1E 6BT, UK.

E-mail: b.gardnersood@ucl.ac.uk

10 weeks. Our experience is that people are reassured to learn that doing the behaviour gets progressively easier; so they only have to maintain their motivation until the habit forms. Working effortfully on a new behaviour for 2–3 months may be an attractive offer if it has a chance of making the behaviour become 'second nature'.

CONCLUSION

Psychological theory and evidence around habit-formation generates recommendations for simple and sustainable behaviour change advice. We acknowledge that health professionals do not always find it appropriate to offer lifestyle counselling to patients: some patients can become annoyed when advised to change their behaviour, and this reaction can threaten patients' trust in and satisfaction with the doctor–patient relationship.[2] However, in settings where professionals feel able to offer behaviour advice, we suggest that they consider providing guidance on habit-formation. Habit-formation advice can be delivered briefly, it is simple for the patient to implement, and it has realistic potential for long-term impact. It offers health professionals a useful tool for incorporating evidence-based health promotion into encounters with patients. A sample tool for health professionals to use with patients to encourage habit formation is provided in Box 1.

Benjamin Gardner,
Lecturer in Health Psychology, Health Behaviour Research Centre, Department of Epidemiology and Public Health, University College London, London.

Phillippa Lally,
ESRC Postdoctoral Research Fellow, Health Behaviour Research Centre, Department of Epidemiology and Public Health, University College London, London.

Jane Wardle,
Professor of Clinical Psychology, Health Behaviour Research Centre, Department of Epidemiology and Public Health, University College London, London.

Provenance
Freely submitted; externally peer reviewed.

DOI: 10.3399/bjgp12X659466

REFERENCES

1. Lansley A. *Response to NHS Future Forum's second report*. London: Department of Health, 2012. http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_132088.pdf (accessed 16 Oct 2012).
2. Lawlor DA, Keen S, Neal RD. Can general practitioners influence the nation's health through a population approach to provision of lifestyle advice? *Br J Gen Pract* 2000; **50(455):** 455–459.
3. Jeffery RW, Drewnowski A, Epstein LH, Stunkard AJ, *et al*. Long-term maintenance of weight loss: current status. *Health Psychol* 2000; **19(Suppl1):** 5–S16.
4. Kahneman D. A perspective on judgment and choice. *Am Psychol* 2003; **58(9):** 697–720.
5. Neal DT, Wood W, Labrecque JS, Lally P. How do habits guide behavior? Perceived and actual triggers of habits in daily life. *J Exp Soc Psychol* 2012; **48:** 492–498.
6. Wood W, Neal DT. A new look at habits and the habit-goal interface. *Psychol Rev* 2007; **114(4):** 843–863.
7. Bayley PJ, Frascino JC, Squire LR. Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature* 2005; **436(7050):** 550–553.
8. Hull CL. *Principles of behavior: an introduction to behavior theory*. New York, NY: Appleton-Century-Crofts, 1943.
9. Lally P, van Jaarsveld CHM, Potts HWW, Wardle J. How are habits formed: modelling habit formation in the real world. *Euro J Soc Psychol* 2010; **40:** 998–1009.
10. Lally P, Wardle J, Gardner B. Experiences of habit formation: a qualitative study. *Psychol Health Med* 2011; **16(4):** 484–489.
11. Gardner B, de Bruijn GJ, Lally P. A systematic review and meta-analysis of applications of the Self-Report Habit Index to nutrition and physical activity behaviours. *Ann Behav Med* 2011; **42(2):** 174–187.
12. Lally P, Gardner B. Promoting habit formation. *Health Psychol Rev*. In press: DOI: 10.1080/17437199.2011.603640.
13. Rothman AJ, Sheeran P, Wood W. Reflective and automatic processes in the initiation and maintenance of dietary change. *Ann Behav Med* 2009; **38(Suppl1):** S4–17.
14. Lally P, Chipperfield A, Wardle J. Healthy habits: Efficacy of simple advice on weight control based on a habit-formation model. *Int J Obes* 2008; **32(4):** 700–707.
15. McGowan L, Cooke LJ, Croker H, *et al*. Habit-formation as a novel theoretical framework for dietary change in pre-schoolers. *Psychol Health* 2012; **27(Suppl1):** 89.
16. Beeken RJ, Croker H, Morris S, *et al*. Study protocol for the 10 Top Tips (10TT) Trial: Randomised controlled trial of habit-based advice for weight control in general practice. *BMC Public Health* 2012; **12(1):** 667.
17. Deci EL, Ryan RM. The support of autonomy and the control of behavior. *J Pers Soc Psychol* 1987; **53(6):** 1024–1037.
18. Gardner B, Lally P. Does intrinsic motivation strengthen physical activity habit? Modeling relationships between self-determination, past behaviour, and habit strength. *J Behav Med* [Epub ahead of print].
19. Hill JO. Can a small-changes approach help address the obesity epidemic? A report of the Joint Task Force of the American Society for Nutrition, Institute of Food Technologists, and International Food Information Council. *Am J Clin Nutr* 2009; **89(2):** 477–484.
20. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. *Can Med J* 2006; **174(6):** 801–809.
21. Bandura A. Social cognitive theory: an agentic perspective. *Annu Rev Psychol* 2001; **52:** 1–26.
22. Maltz M. *Psycho-cybernetics*. New York, NY: Prentice Hall, 1960.

THIS ARTICLE HAS BEEN CORRECTED. SEE LAST PAGE

Developmental Psychology
2005, Vol. 41, No. 4, 625–635

© 2005 American Psychological Association
0012-1649/05/$12.00 DOI: 10.1037/0012-1649.41.4.625

Peer Influence on Risk Taking, Risk Preference, and Risky Decision Making in Adolescence and Adulthood: An Experimental Study

Margo Gardner and Laurence Steinberg
Temple University

In this study, 306 individuals in 3 age groups—adolescents (13–16), youths (18–22), and adults (24 and older)—completed 2 questionnaire measures assessing risk preference and risky decision making, and 1 behavioral task measuring risk taking. Participants in each age group were randomly assigned to complete the measures either alone or with 2 same-aged peers. Analyses indicated that (a) risk taking and risky decision making decreased with age; (b) participants took more risks, focused more on the benefits than the costs of risky behavior, and made riskier decisions when in peer groups than alone; and (c) peer effects on risk taking and risky decision making were stronger among adolescents and youths than adults. These findings support the idea that adolescents are more inclined toward risky behavior and risky decision making than are adults and that peer influence plays an important role in explaining risky behavior during adolescence.

Keywords: adolescents, risk taking, peer influence, risk preference, decision making

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

It is well documented that adolescents are more likely than adults to engage in risky behavior. For example, adolescents are more likely than adults to drive recklessly, to drive while intoxicated, to use varied illicit substances, to have unprotected sex, and to engage in both minor and more serious antisocial behavior (Arnett, 1992). However, despite clinical and anecdotal evidence of heightened real-world risk taking during adolescence, laboratory studies of age differences in risk preference, risk perception, and risky decision making have not yielded consistent evidence that adolescents are actually less risk averse than are their elders. In fact, it is often asserted that, by midadolescence, teens' capacities for understanding and reasoning in risky decision-making situations roughly approximate those of adults (Fischhoff, 1992; Furby & Beyth-Marom, 1992). This assertion has been used to argue both for protecting adolescents' rights to make autonomous decisions about their reproductive health and for holding adolescents to adult standards of criminal blameworthiness (see Steinberg & Scott, 2003, for a discussion).

However, as several writers have recently argued, extant studies suggesting equivalent orientations toward risk among adolescents and adults are only modestly useful in understanding how adolescents compare with adults in real-world decision making. These authors suggest that typical laboratory studies of risky decision making fail to consider the emotional and social contexts in which risk taking actually occurs (Cauffman & Steinberg, 2000; Scott, Reppucci, & Woolard, 1995; Steinberg, 2004; Steinberg & Cauffman, 1996). In such studies, individual adolescents are presented with hypothetical dilemmas under conditions of low emotional arousal and are then asked to make and explain their decisions. In the real world, however, adolescents' decisions are not hypothetical, they are generally made under conditions of emotional arousal (whether negative or positive), and they are usually made in peer groups. Whether the risky decision making of adolescents is truly comparable to that of adults under real-world conditions remains an open and unstudied question.

A number of explanations have been advanced to account for differences between adolescents and adults in real-world, as opposed to laboratory-based, risk taking. Some have argued that age differences in psychosocial capacities such as impulse control or sensation seeking play an important role (see Steinberg & Cauffman, 1996). Consistent with this, Cauffman and Steinberg (2000) reported that once differences in psychosocial maturity between adolescents and adults are accounted for, age differences in risky decision making disappear. An alternative and entirely compatible account of age differences in risky behavior emphasizes the role of peers and, more specifically, peer influence. That is, adolescents may engage in more risky behavior than do adults because they are more susceptible to the influence of their similarly risk-prone peers. Support for this latter explanation comes, in part, from the criminology literature. There is a small but compelling body of evidence to suggest that when adolescents commit crimes—acts that are inherently risky—they generally do so with their peers (Erickson & Jensen, 1977; Zimring, 1998). For example, adolescents are usually accompanied by one or more persons when committing crimes that range in seriousness from vandalism and drug use (Erickson & Jensen, 1977) to rape and homicide (Zim-

Margo Gardner and Laurence Steinberg, Department of Psychology, Temple University.

This study was supported by a grant from the John D. and Catherine T. MacArthur Foundation Network on Adolescent Development and Juvenile Justice. We thank Rebecca Davis, Rebecca Garrett, Lauren Guttshall, Nermine Salama, Benjamin Steinberg, and Erica Weitz for help with data collection. We also thank He Len Chung, Alex Piquero, and Jennifer Silk for their help with the selection and implementation of the analytic method used.

Correspondence concerning this article should be addressed to Margo Gardner at the Department of Psychology, Temple University, Philadelphia, PA 19122. E-mail: mnoel002@temple.edu

ring, 1998). This is not, however, true of adults; when adults commit crimes, they typically do so alone (Zimring, 1998).

Although adolescent risk taking often occurs in groups, it is not known whether the greater prevalence of group risk taking observed among adolescents stems from the fact that adolescents spend more time in peer groups than adults do (Brown, 2004) or from the heightened levels of susceptibility to peer influence that have been shown to characterize adolescence (Steinberg & Silverberg, 1986). In other words, it is not clear whether adolescents simply have more opportunities to engage in group risk taking than do adults or whether, when faced with behavioral decisions in a peer group context, adolescents are more easily swayed toward risky choices.

To our knowledge, only one study has attempted to determine whether there are developmental differences in the effects of actual peer presence on orientation toward risk. In a comparison of adolescents and college students, Hensley (1977) sought to determine whether the tendency for individuals to take more risks in groups than when alone—a phenomenon known as the *risky shift* (Vinokur, 1971)—might differ across age groups. Hensley (1977) used a hypothetical decision-making questionnaire to measure risk acceptance and found that the magnitude of the risky shift was greater among adolescents than it was among college students. However, the study sample was very small (22 college students and 18 adolescents), and these results have not, to our knowledge, been replicated with performance (as opposed to hypothetical) measures of risk taking.

There are other findings that indirectly support the notion that adolescents may be more easily swayed toward risky behavior than adults. Compared with adults, adolescents have limited abilities in areas of psychosocial functioning, such as self-reliance, which likely interfere with the ability to act independently of the influence of others (Cauffman, 1996; Cauffman & Steinberg, 2000; Steinberg & Cauffman, 1996). Not surprisingly, several studies have found a curvilinear relation between age and peer conformity on responses to hypothetical dilemmas about antisocial decision making, with conformity increasing throughout childhood and into midadolescence and decreasing thereafter (Berndt, 1979; Brown, Clasen, & Eicher, 1986; Steinberg & Silverberg, 1986). Although researchers have not examined the developmental pattern of resistance to peer influence beyond late adolescence, there is some evidence that peer influence remains an important predictor of participation in risky behavior even during young adulthood (Andrews, Tildesley, Hopps, & Li, 2002; Horvath & Zuckerman, 1993). Thus, when confronted with risky decisions in the context of a peer group, adolescents, and perhaps even young adults, may be less able than older adults to resist the influence of their risk-prone age mates.

Further support for the idea of heightened peer effects on risky behavior during adolescence comes from additional findings on the risky-shift. Although a number of researchers have found that risk-taking tendencies are greater when individuals are in groups than when alone (e.g., Blascovich & Ginsburg, 1974; Blascovich, Ginsburg, & Howe, 1975; Blascovich, Veach, & Ginsburg, 1973; Kogan & Wallach, 1967; Lamm, 1967; Lamm, Trommsdorff, & Rost-Schaude, 1972; Pruitt & Teger, 1969; Vidmar, 1970; Wallach & Kogan, 1965; Yinon, Jaffe, & Feshbach, 1975), several investigations have found the reverse to be true. Indeed, in some cases, individuals demonstrate a conservative shift and are actually more risk averse when in groups than when alone (e.g., Cohen & Ruis, 1974; Pilkonis & Zanna, 1973; Zaleska, 1974). Accordingly, social psychologists have advanced an alternative theoretical framework for understanding group risk taking. Whereas proponents of the risky shift theory assert that the presence of others should always lead to increased risk taking, advocates for the more recent group polarization theory suggest that the direction of group effects on risk taking depends on the risk-taking tendencies of the group members (Hogg, Turner, & Davidson, 1990). According to this theory, relatively conservative individuals should become even more conservative when grouped together, whereas individuals who are inclined to take risks should make even more risky choices (Hogg et al., 1990). Given this theoretical framework, adolescents' generally greater inclination toward risky behavior as individuals, in combination with their greater susceptibility to peer influence should, in theory, result in a larger effect of peer presence on risky behavior among adolescents than among adults.[1] The goal of the present study is therefore to examine whether adolescents, relative to adults, are more likely to take risks when their peers are present.

The Present Study

In the present study, we examined the differential effects of the presence of peers on risk taking, risk preference, and risky decision making among adolescents (*M* age = 14), youths (*M* age = 19), and adults (*M* age = 37). Our three primary hypotheses were as follows:

Hypothesis 1. Risk taking, risk preference, and risky decision making will decrease with age.

Hypothesis 2. On average, individuals will demonstrate more risk taking, greater risk preference, and more risky decision making when in the company of their peers than when alone.

Hypothesis 3. The difference between levels of risk taking, risk preference, and risky decision making with and without the presence of peers will decrease with age. That is, group effects on risk orientation will be greater among adolescents than among youths, and greater among youths than among adults.

Method

Sample

Our sample included 106 adolescents (54 girls and 52 boys), ages 13 to 16 (*M* age = 14.01, *SD* = 1.02), 105 youths (53 women and 52 men), ages 18 to 22 (*M* age = 18.78, *SD* = 1.07), and 95 adults (48 women and 47 men), ages 24 and older (*M* age = 37.24, *SD* = 12.37). All participants were recruited from areas in and around a major urban center. The adolescents were recruited from middle schools, day camps, and commu-

[1] In keeping with group polarization theory, we are not suggesting that all adolescents should demonstrate shifts toward increased risk taking when in the presence of peers. It is conceivable that some adolescents, when placed in a group with risk-averse peers, might shift toward decreased risk taking. However, we expect that, given generally greater propensities for risk taking among adolescents, adolescents should, on average, be more likely than adults to demonstrate group-induced shifts toward greater risk taking.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

nity centers; the youths were recruited from undergraduate introductory psychology courses at a large urban university; and the adults were recruited through fliers posted on urban university and community college campuses, advertisements distributed to community organizations, and word of mouth.

The adolescent sample was composed of 50.9% girls and 49.1% boys; the youth sample was composed of 50.5% women and 49.5% men; and the adult sample was composed of 50.5% women and 49.5% men. These three groups did not differ significantly with respect to gender composition, $\chi^2(2) = .005$, $p = .997$. The three age groups were also very similar in terms of their ethnic composition. The majority of the participants were either White (48.7%) or African American (38.2%). Given the very small percentage of participants from other ethnic groups (the sample included only 1% Native Americans, 7.2% Asian Americans, 3.9% Latinos, and 0.7% others), the three age groups were compared only with respect to the percentages of White versus non-White participants. The adolescent sample included 44.8% White participants and 55.2% non-White participants; the youth sample included 53.3% White participants and 46.7% non-White participants; and the adult sample included 48.4% White participants and 51.6% non-White participants. These three groups did not differ significantly with respect to ethnic composition, $\chi^2(2) = 1.554$, $p = .460$. Finally, on a participant-report parent education rating scale (a proxy for socioeconomic status) where 1 = *less than high school diploma*, 2 = *high school diploma/GED*, 3 = *some college/vocational school*, 4 = *college graduate*, and 5 = *graduate or professional school*, all three age groups reported a mean level of parent education within the range of some college or vocational training. Thus, the three age groups did not differ substantially with respect to gender, ethnicity, or socioeconomic status (as indexed by parent education).

Recruitment and Procedure

Slightly different recruitment procedures were used for the adolescents versus the college students and adults. Each college undergraduate and adult participant was asked to invite to the session 2 people of his or her gender that he or she knew. These groups of 3 were then randomly assigned to a group condition (in which all 3 participants completed the battery of measures at the same time, in the same room, and communicated with each other while completing the tasks and measures) or to a sole participant condition (in which each of the 3 participants completed the battery of measures alone while the other 2 participants waited outside the room where the session took place).

Requirements concerning the need for parental consent among the adolescent participants necessitated a slight modification in this procedure. We found it difficult to get predetermined groups of 3 adolescents to appear for scheduled appointments together and to arrive with all 3 signed parent consent forms in hand. Thus, after first obtaining parental consent from large numbers of adolescents at different recruitment sites, we randomly assigned half of these adolescents within each recruitment site to the group condition (in which they completed the measures with two other teenagers of the same gender from their recruitment site) and half to the sole participant condition (in which they completed the measures alone). Thus, while the adolescent triads in the group condition were always composed of individuals who knew and were familiar with each other (i.e., individuals within each triad were from the same camp, classroom, or community-center summer or afterschool program), they were not necessarily composed of individuals who had selected one another as partners for the experiment. Although the adolescent and two older groups differed in this respect, in the real world, adolescents often find themselves in groups with other teenagers whom they know but have not necessarily selected as companions (i.e., in classrooms, on sports teams, in extracurricular activities, etc.). Moreover, to the extent that friends have been shown to exert more pressure on each other than acquaintances (e.g., McPhee, 1996), the way in which adolescents were recruited resulted in a more conservative test of the hypothesis that, relative to adults, risk orientation among adolescents is more influenced by peer pressure.

Among the adults, 54 participants (27 men and 27 women) were assigned to the group condition, and 41 participants (20 men and 21 women) were assigned to the sole participant condition.[2] Among the youths, 54 participants (27 men and 27 women) were assigned to the group condition, and 51 (25 men and 26 women)[3] were assigned to the sole participant condition. Among the adolescents, 54 participants (27 boys and 27 girls) were assigned to the group condition, and 52 (25 boys and 27 girls) were assigned to the sole participant condition.

All participant triads were composed of individuals of the same gender. However, the ethnic composition of the triads was not constricted in this manner. Among the adults, 37.5% of the triads consisted of all White participants, 43.8% consisted of all non-White participants, and 18.8% of the triads consisted of White and non-White individuals. Among the youths, 45.9% of the triads consisted of all White participants, 37.8% consisted of all non-White participants, and 16.2% consisted of White and non-White participants. Among the adolescents, 19.4% of the triads consisted of all White participants, 25% consisted of all non-White participants, and 55.6% consisted of White and non-White participants.

All participants completed three measures of risk orientation that were part of a battery of measures administered for a larger study of psychosocial development. The entire battery of measures took approximately 1 hr to complete. Each adolescent and adult participant was compensated $20, and each undergraduate participant was given the choice between either a $20 payment or research credit in an introductory psychology course.

Measures

Risk taking. Risk taking was assessed with a video game called "Chicken" (Sheldrick, 2004). Chicken is played on a laptop computer and requires participants to make decisions about whether to stop a car that is moving across the screen once a traffic light turns from green to yellow. The appearance of the yellow light signals the impending appearance of a red traffic light, as well as a potential crash if the car is still moving when the red light appears. Chicken was selected because it measures risk taking in the moment rather than the more deliberative form of risk taking assessed in many studies, in which participants have unlimited time to consider and evaluate all potential decisions and outcomes. Additionally, Chicken requires participants to make actual decisions in a risky situation, rather than simply requiring participants to report what they would do in a hypothetical risky situation.

The game is played from a third-person, side-view perspective (see Figure 1) and consists of 15 trials. In each trial, participants watched an animated car move across the screen for a predetermined amount of time until a yellow traffic light appeared. Before the first trial, players were informed that at some unknown point after the yellow light appeared, the traffic light would turn red and a wall would pop up in front of the car. Players were told that the object of the game was to allow the car to move as far as possible without crashing into the wall. Players controlled whether the car was moving or stopped but not the speed of the car. Participants accumulated more points the further the car moved without crashing but lost any points that had been accumulated on that trial if the

[2] One person was dropped from one of the adult sole participant triads. At the age of 79, we believed this individual to be developmentally dissimilar from the rest of the adult sample.

[3] One person was missing from three of the young adult triads assigned to the sole participant condition. In the first two cases, the 3rd member of the group left before it was his or her turn to complete the session. In the third case, data for the 3rd member of the peer group were excluded because it was later determined that the participant had already completed the study several months prior.

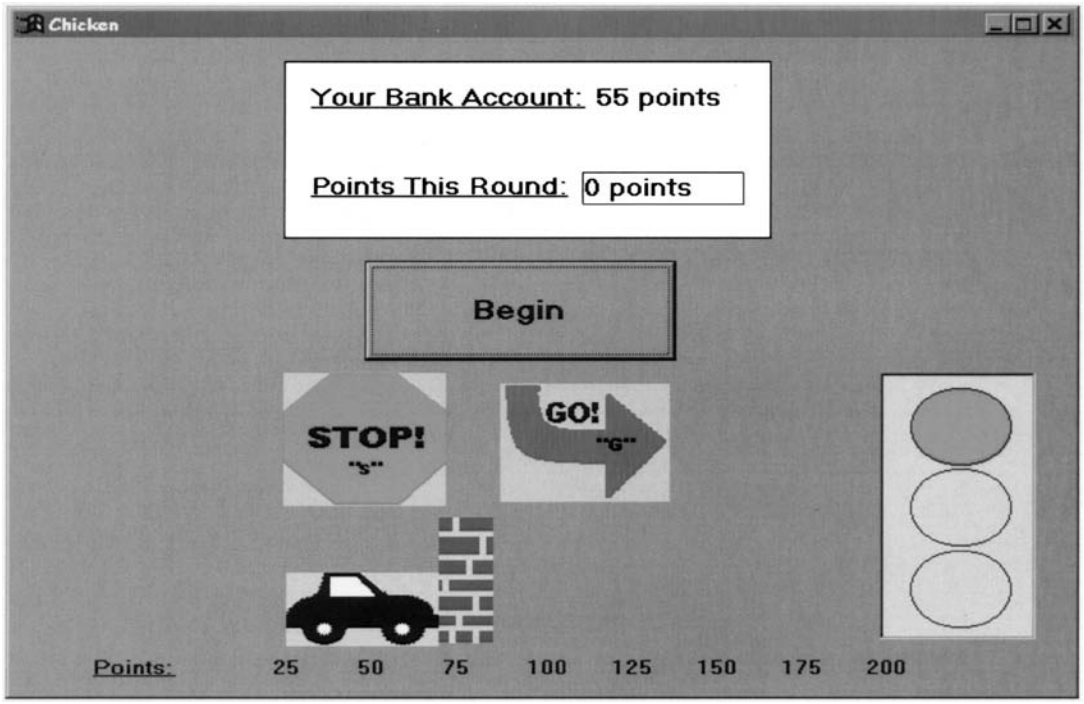


Figure 1. An image from the Chicken video game. In this frame, the traffic light has just turned red. The car was still moving when the light turned red; consequently, a brick wall appeared in front of the car, resulting in a crash.

car crashed. If a player stopped the car before it crashed, the player had the option of restarting the car and allowing it to move further, or leaving the car where it was and accepting the amount of points accumulated. Thus, when the yellow light appeared, players had to decide how much further to allow the car to move, balancing their desire to accumulate points against the possibility of crashing the car into the wall. The latency between the beginning of the trial and the appearance of the yellow light, and between the appearance of the yellow light and the appearance of the wall varied across trials, such that the participants did not know whether the wall would appear suddenly or after some delay. The objectives of the game and the potential positive and negative outcomes (earning points vs. crashing and failing to accumulate points, respectively) were explained to participants during a demonstration. In order to ensure that all participants were equally familiar with the potential consequences of driving through a yellow traffic light, the demonstration round included a depiction of the animated car both driving safely through the yellow light without crashing, as well as driving through the yellow traffic light and crashing into the pop up wall.

The computer recorded the amount of time that the car was in motion between the onset of the yellow light and the car's final stop, as well as the number of car restarts per round. Mean scores for the number of car restarts per round, and the percentage of time the car was in motion were calculated for each participant. Longer moving times and more restarts indicated greater risk taking. Scores on these two indices of risk taking were highly correlated ($r = .61$, $p < 01$) and were therefore standardized and averaged in order to compute a composite indicator of risk taking on the Chicken game.

Participants in the sole participant condition completed the task as described. Participants in the group condition took turns playing the game, but all of them completed 15 trials in a row, as did participants in the sole participant condition. In the group condition, while one participant was playing the game, the other two were told that they could call out advice about whether to allow the car to keep moving or to stop it. The player was instructed that he or she could choose whether to follow the advice of his or her peers.

Risk preference.

A shortened, modified version of the Benthin Risk Perception Measure (BRPM; Benthin, Slovic, & Severson, 1993) was used to assess risk preference. This measure assesses both risk perception (the extent to which one perceives a given activity as carrying the potential for adverse consequences) and risk preference (whether one believes the benefits inherent in an activity outweigh the costs, or vice versa). Only data from the scale reflecting cost–benefit consideration are used in the present analyses. We chose not to include data from the risk perception scale because prior studies have failed to find age differences in performance on this scale (Steinberg, 2004). Similarly, evidence from the research of Beyth-Marom and colleagues (Beyth-Marom, Austin, Fischoff, Palmgren, & Jacobs-Quadrel, 1993) suggests that adolescents and adults are relatively equal in terms of their awareness of the potential for adverse consequences in risky situations. However, as argued by Furby and Beyth-Marom (1992), adolescents and adults may differ in terms of the relative weights or values that they attach to the potential costs and benefits of risky activities. Studies of the relation between risk taking and cost versus benefit consideration suggest that those who give lesser consideration to costs

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

and greater consideration to benefits are more likely to engage in risky behavior (e.g., Fromme, Stroot, & Kaplan, 1993; Goldberg & Fischhoff, 2000; Horvath & Zuckerman, 1993; Lavery, Siegel, Cousins, & Rubovits, 1993; McBride, Weatherby, Inciardi, & Gillespie, 1999; Singer, Dai, Weeks, & Malave, 1998; Thorton, Gibbons, & Gerrard, 2002). Thus, differential consideration or weighting of potential costs versus benefits among adolescents and adults may partially account for observed age differences in risky behavior.

In completing the Risk Preference Scale, participants were presented with five hypothetical scenarios involving risky behavior. These scenarios included having sex without a condom, riding in a car driven by someone who has been drinking, trying a new drug that one does not know anything about, breaking into a store at night and stealing something that one really wants, and driving over 90 mph on the highway at night. They were then asked to rate on a 4-point scale ranging from 1 (*risks are much greater than benefits*) to 4 (*benefits are much greater than risks*) how the risks compared with the benefits of the activity. A mean risk–benefit consideration score was then calculated for each participant by averaging responses across the five scenarios ($\alpha = .68$).

Individuals in the sole participant condition read the scenarios from index cards and indicated their choices on a response card displaying the 4-point scale. Group condition participants followed the same procedure but were told that they could discuss each question. However, they were instructed that they need not reach a consensus and that each could make a final decision at any time. Each participant had his or her own set of response cards and had an unobstructed view of the others' response cards. The administrator recorded individuals' responses.

Risky decision making.

Risky decision making was assessed via the Youth Decision-Making Questionnaire (YDMQ; Ford, Wentzel, Wood, Stevens, & Siesfeld, 1990). Participants were presented with five hypothetical dilemmas, each involving a risky decision. The dilemmas included decisions about allowing friends to bring drugs into one's home, stealing a car, cheating on an exam, shoplifting, and skipping work without an excuse, all of which adolescents, college undergraduates, and adults potentially could have done. Decisions about each dilemma were made within the context of three different scenarios. In the first scenario, participants were informed that no matter what their decision, no negative consequences would result. The second scenario—introduced by Cauffman and Steinberg (2000)—stated that negative consequences might result if the risky course of action were taken. The final scenario stated that negative consequences would definitely occur if the risky course of action were taken.

Only responses from the second decision-making scenario (i.e., negative consequences might result) were included in the analyses for the present study, as this was the only scenario that involved some degree of uncertainty or risk. For each dilemma, participants were asked to decide what they would do "if they were really in that situation" on a 4-point scale that ranged from 1 (*definitely making the risky decision*) to 4 (*definitely not making the risky decision*). Scores were reverse coded, such that higher scores indicated higher risk-taking tendencies. A mean risky decision-making score was calculated for each participant by averaging the scores across the five dilemmas ($\alpha = .65$).

Table 1
Intercorrelations, Means, and Standard Deviations of Study Variables

Variable	1	2	3	4	*M*	*SD*
1. Chronological age	—	−.243**	−.279**	−.091	22.77	11.98
2. Risk taking (Chicken)		—	.147*	.127*	0.00	0.90
3. Risky decision making (YDMQ)			—	.331**	2.01	0.56
4. Risk preference (BRPM)				—	1.50	0.50

Note. YDMQ = Youth Decision-Making Questionnaire; BRPM = Benthin Risk Perception Measure.
* $p < .05$. ** $p < .01$.

The YDMQ was presented on a laptop computer. Individuals in the sole participant condition were asked to indicate their desired choices on cards displaying the 4-point response scale. Participants in the group condition followed the same procedure but were told that they could discuss each situation. They were also informed that they did not need to reach a consensus and that they could each make a final decision at any time. Each group condition participant had his or her own set of response cards, and each had an unobstructed view of the others' cards. The administrator recorded participants' responses. Means, standard deviations, and intercorrelations among the study variables are presented in Table 1.

Results

Data Analyses

Because participants were recruited in groups of 3, scores for participants within each triad could not be treated as independent.[4] In order to accommodate the nested structure of the data, all analyses were performed with the linear mixed model (LMM) procedure in the Statistical Package for Social Sciences 11.5 (SPSS; SPSS, Inc., 2005). Unlike the general linear model (GLM) procedure, which assumes that all observations are independent of one another, the LMM procedure allows for correlated variability among observations. Because the LMM procedure does not permit the simultaneous analysis of multiple dependent variables, separate LMM analyses were performed for each of the three dependent variables (Chicken, BRPM, YDMQ). Prior to entering the independent and dependent variables for each analysis, the structure of the data—individuals nested within triads—was specified. Then, for each analysis, chronological age was entered as a continuous independent variable, and condition (group vs. sole participant)

[4] The adolescent sole participants were not recruited in groups of 3. However, in order to structure the data from the three age groups as similarly as possible, triads of adolescent sole participants were created for purposes of data analyses. The adolescent sole participant sample was subdivided by data collection site and then further subdivided by gender, such that a female from a particular community center could only be grouped with another female from that same community center, or a male from a particular middle school could only be grouped with another male from that same middle school. This was done under the assumption that adolescents of the same gender from the same site would most likely know one another, thus making the triads of adolescent

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

was entered as a fixed factor. Additionally, gender and ethnicity (White vs. non-White) were entered as fixed variables in order to determine whether these variables moderated age, condition, or Age × Condition effects.

Age differences in risk taking, risk preference, and risky decision making. The effect of chronological age on risk taking and risky decision making was significant, $F(1, 284) = 18.79$, $p < .0001$, $r_{\text{effect size}} = .249$, and, $F(1, 288) = 24.599$, $p < .0001$, $r_{\text{effect size}} = .281$, respectively. During the risk-taking game, younger individuals allowed the car to move forward for longer periods of time after the appearance of the yellow light and were more likely to restart the car after stopping it. Similarly, younger individuals were more likely than older participants to select the risky course of action on the risky decision-making questionnaire. The effect of chronological age on risk preference was not significant, however, $F(1, 288) = .563$, $p = .465$.

Effect of peer presence on risk taking, risk preference, and risky decision making. We found significant effects of peer presence on all three measures of risk orientation. Specifically, compared with those who completed the measures by themselves, participants who completed the same measures with peers present took more risks during the risk-taking game, $F(1, 284) = 15.05$, $p < .0001$, $r_{\text{effect size}} = .224$; gave greater weight to the benefits rather than the costs of risky activities, $F(1, 288) = 3.662$, $p = .057$, $r_{\text{effect size}} = .112$; and were more likely to select risky courses of action in the risky decision-making situations, $F(1, 288) = 6.308$, $p < .05$, $r_{\text{effect size}} = .146$.

Differential effects of peer presence on risk taking, risk preference, and risky decision making as a function of age. The effects of peer presence varied as a function of age on the risk-taking measure, $F(1, 284) = 4.801$, $p < .05$, $r_{\text{effect size}} = .129$, and the risky decision-making measure, $F(1, 288) = 4.943$, $p < .05$, $r_{\text{effect size}} = .130$, but not on the Risk Preference Scale, $F(1, 293) = .284$, $p = .594$. As Figure 2 indicates, for example, the magnitude of the group effect on risk taking was greater among younger rather than older participants (see Table 2 for means and standard deviations). The pattern of results was similar with respect to risky decision making.

The Effects of Gender and Ethnicity

The effects of gender and ethnicity on risk orientation were not focal issues in the present study. Thus, no hypotheses on the effects of these variables were generated. Nonetheless, gender and ethnicity were included in the model in order to determine whether the age, condition, or Age × Condition interaction effects differed across males and females, or between White and non-White individuals.

We found few significant gender effects. There were no differences between males and females on risk taking or risky decision making, nor were there any significant two-way interaction effects involving gender on measures of these constructs. Additionally,

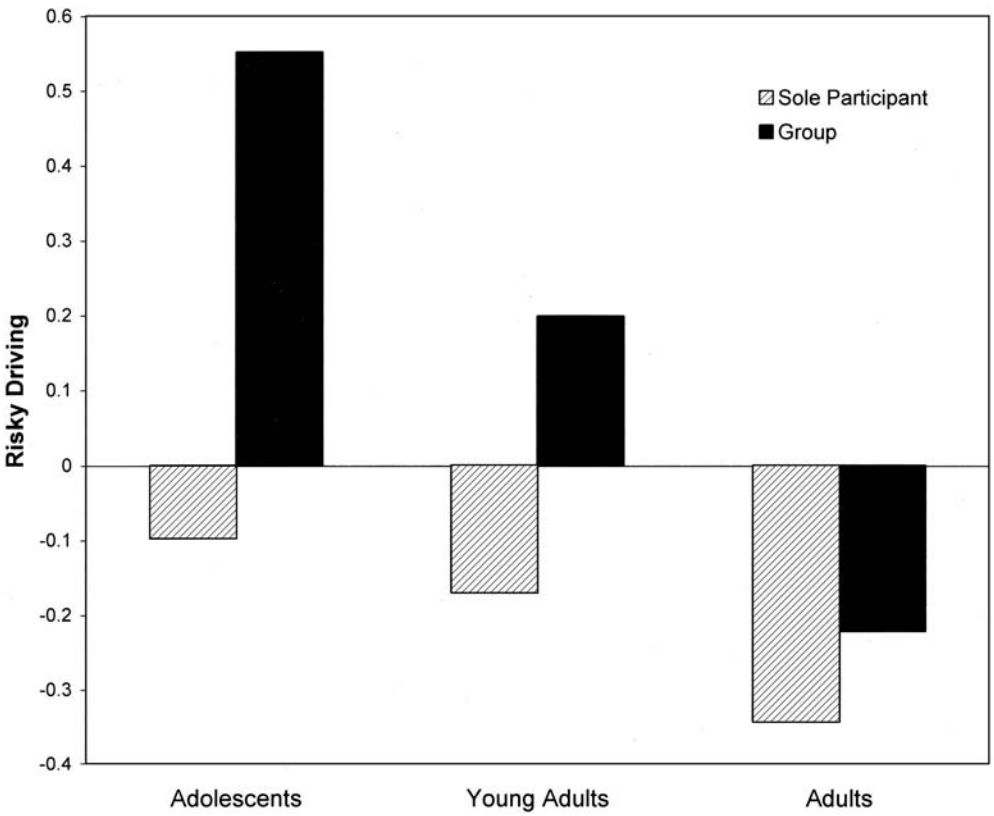


Figure 2. Age × Condition interaction on Chicken game, where higher scores indicate more risk taking.

sole participants as similar to those of undergraduate and adult sole participant triads as possible.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
Descriptive Statistics for Group and Sole Participant Conditions by Age Group

		White			Non-White			Total		
Age	Condition	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*
					Risk taking (Chicken)					
Adolescent	Sole	−.164	.612	22	−.035	.722	29	−.097	.669	52
	Group	.140	.886	25	.907	1.300	29	.552	1.182	54
Youth	Sole	−.258	.729	24	−.091	.666	27	−.170	.694	51
	Group	.139	.848	30	.289	.893	20	.199	.860	50
Adult	Sole	−.367	.387	22	−.316	.620	19	−.343	.502	41
	Group	−.080	.566	24	−.335	1.167	30	−.221	.949	54
					Risky decision making (YDMQ)					
Adolescent	Sole	2.127	.511	22	1.862	.571	29	1.977	.551	52
	Group	2.072	.519	25	2.021	.522	29	2.044	.516	54
Youth	Sole	2.142	.564	24	2.074	.448	27	2.106	.502	51
	Group	2.506	.477	32	2.282	.358	22	2.428	.455	50
Adult	Sole	1.655	.339	22	1.962	.758	19	1.781	.581	41
	Group	1.625	.397	24	1.720	.497	30	1.678	.453	54
					Risk preference (BRPM)					
Adolescent	Sole	1.391	.335	22	1.372	.286	29	1.385	.304	52
	Group	2.072	.519	25	2.021	.522	29	1.593	.764	54
Youth	Sole	2.142	.564	24	2.074	.448	27	1.451	.447	51
	Group	1.819	.600	32	1.555	.365	22	1.724	.536	50
Adult	Sole	1.427	.317	22	1.368	.354	19	1.400	.332	41
	Group	1.425	.355	24	1.393	.358	30	1.407	.354	54

Note. Chicken means are based on standardized scores. YDMQ = Youth Decision-Making Questionnaire; BRPM = Benthin Risk Perception Measure.

we failed to find significant Age × Condition × Gender interactions on any measure of risk orientation. Nevertheless, we did find main effects of gender and gender differences in age and condition effects on the measure of risk preference. First, males gave significantly greater weight to the benefits of risky decisions than did females, $F(1, 288) = 19.961$, $p < .0001$, $r_{\text{effect size}} = .255$. Second, we found that males weighted the benefits of risky activities more heavily when in a group than when alone, but that cost–benefit consideration did not differ substantially between the group and sole participant conditions among females, $F(1, 288) = 6.058$, $p < .05$, $r_{\text{effect size}} = .144$. Finally, we found that among younger individuals, males weighted the benefits of risky decisions more heavily than did females but that among older individuals males and females gave comparable weights to the benefits of risky decisions, $F(1, 288) = 11.089$, $p < .01$, $r_{\text{effect size}} = .193$.

In contrast to these limited gender differences, a number of significant ethnicity effects were identified. First, we found significant differences between White and non-White participants on the measures of risk taking, $F(1, 284) = 11.67$, $p < .01$, $r_{\text{effect size}} = .199$, and risky decision making, $F(1, 288) = 6.645$, $p < .01$, $r_{\text{effect size}} = .150$. However, the direction of these effects differed. Although non-White participants engaged in greater risk taking than did White participants, White participants made more risky decisions than did non-White participants.

Second, the effects of age on risk taking, risky decision making, and risk preference differed across White and non-White individuals, $F(1, 284) = 9.03$, $p < .01$, $r_{\text{effect size}} = .176$; $F(1, 288) = 4.289$, $p < .05$, $r_{\text{effect size}} = .121$; and, $F(1, 288) = 3.922$, $p < .05$, $r_{\text{effect size}} = .116$, respectively (see Table 2). Overall, as noted earlier, adolescents and youths were more oriented toward risk than were adults (although the age effect on risk preference measure did not reach significance). However, within the age groups, risk taking, risk preference, and risky decision making varied somewhat as a function of ethnicity. For instance, although there were negligible differences between White and non-White adults in risk taking and risk preference, among adolescents non-White individuals took more risks and demonstrated a greater preference for risk than did White individuals. Conversely, whereas non-White adults were slightly more likely to make risky decisions than were White adults, White adolescents were slightly more likely to make risky decisions than were non-White adolescents.

Third, White and non-White participants differed in their response to peer presence on the measures of risk taking, $F(1, 284) = 4.383$, $p < .05$, $r_{\text{effect size}} = .123$, and risk preference, $F(1, 288) = 6.517$, $p < .05$, $r_{\text{effect size}} = .149$ (see Table 2). Peer presence was associated with greater risk taking and risk preference among both White and non-White participants. However, condition (group vs. sole participant) differences in mean risk-taking scores were greater among non-White than were those among White participants. On the risk-preference measure, effect sizes for condition differences mirrored this pattern ($r_{\text{effect sizes}} = .203$ and .108 for non-White and White participants, respectively).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

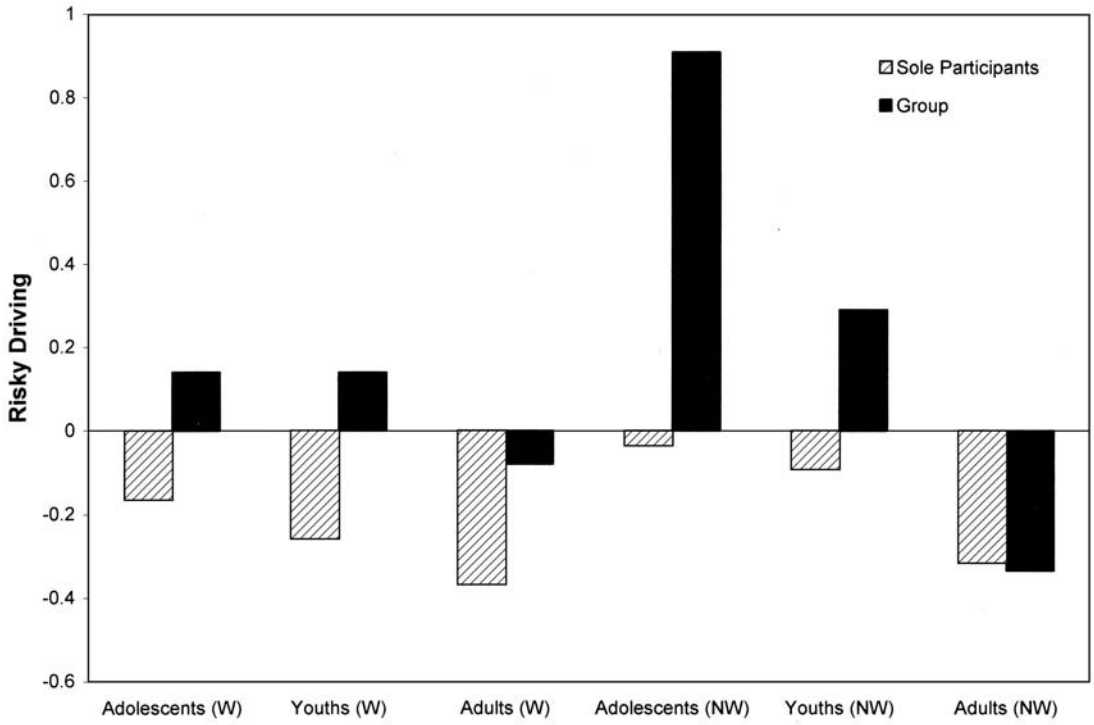


Figure 3. Age × Condition × Ethnicity Interaction on Chicken game, where higher scores indicate more risk taking. W= White participants; NW = non-White participants.

However, effect size estimates indicated the opposite pattern for risk taking ($r_{\text{effect sizes}}$ = .196 and .233 for non-White and White participants, respectively). Effect size estimates are reduced by variability, and although condition differences in mean risk taking were greater among non-White than White participants, scores among non-White participants were more variable (see Table 2 for standard deviations).

Finally, we found significant Age × Condition × Ethnicity interaction effects on risk taking, $F(1, 284) = 4.011$, $p < .05$, $r_{\text{effect size}} = .118$, and risk preference, $F(1, 288) = 5.961$, $p < .05$, $r_{\text{effect size}} = .142$; see Table 2). As noted earlier, peer presence had a greater impact on risk orientation among adolescents and youths than among adults (although the two-way Age × Condition interaction for risk preference did not reach significance). However, the effects of peer presence on the risk-taking and risk-preference tendencies of individuals within some, but not all, of the age groups varied as a function of ethnicity. For instance, among adults, there were no differences between White and non-White participants in the effects of peer presence on risk preference. However, among adolescents, peer effects on risk preference were greater for non-White than for White participants. A similar pattern emerged on the risk-taking measure. That is, peer effects on risk taking were greater among non-White than among White adolescents. However, on this measure, peer effects on adult risk-taking tendencies were greater for White than for non-White participants (see Figure 3).

Discussion

Between adolescence and adulthood there is a significant decline in both risk taking and risky decision making. In addition, our findings suggest that, in some situations, individuals may take more risks, evaluate risky behavior more positively, and make more risky decisions when they are with their peers than when they are by themselves. Most importantly, the effects of peer presence on both risk taking and risky decision making vary as a function of age. That is, although the sample as a whole took more risks and made more risky decisions in groups than when alone, this effect was more pronounced during middle and late adolescence than during adulthood. Thus, relative to adults, adolescents are more susceptible to the influence of their peers in risky situations.

The methodological strengths of this study provide good reason to feel confident about the internal validity of the findings. First, as an experiment that uses random assignment, we were able to control individuals' exposure to peers. Second, whereas the only previous developmental comparison of peer effects on risk taking (Hensley, 1977) relied on a hypothetical decision-making questionnaire and used risk acceptance as a proxy for risk taking, the battery of measures in the present study included not only hypothetical decision-making questionnaires but also a behavioral measure of risk taking that required participants to make actual decisions about how much risk to take in a situation that closely mirrors one faced in everyday life—whether to "run" a yellow light and continue through an intersection. Third, the use of friends, or at least familiar individuals (in the case of the adoles-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

cents), helped to create a more ecologically valid social context than those of many studies of group behavior (e.g., Vidmar, 1970; Wallach & Kogan, 1965; Yinon et al., 1975). Everyday group decision-making situations generally involve friends or acquaintances, and laboratory studies that do not use such groups may not capture the dynamics of real-life group decision making.

It is also necessary to recognize several of the study's limitations. First, although the driving game Chicken is a closer approximation to real-life risk-taking situations than are the typical decision-making questionnaires used in most research of this sort, no laboratory task can adequately simulate real life. No matter how realistic the task, it is difficult to determine whether participants' performance in the laboratory is an accurate representation of their real-world behavior.

Second, different recruitment procedures were used for the adolescents versus the youths and adults. The youths and adults came to the sessions in groups of 3 friends, but the adolescents were assigned to groups of 3 (although all groups of 3 were made up of individuals who were from the same classroom, camp, or community-center program, and who were familiar with one another). Thus, it is conceivable that the older individuals knew one another better than did the adolescents. However, we believe that differences in familiarity among the age groups were minimal and that any effects of peer familiarity of behavior resulted in a more conservative test of our central hypothesis (i.e., that, relative to adults, adolescents should be more easily swayed toward risky behavior by their friends). Although some contradictory findings do exist (e.g., Leary et al., 1994), there is evidence to suggest that the effects of peer familiarity on behavior may either be negligible, or may be stronger when in the company of friends versus acquaintances or strangers. For example, in a study of impression management among young adults, Bohra and Pandey (1984) found very few differences in the attempts of participants to manage the impressions of friends versus strangers. Moreover, in the few cases in which differences were found (e.g., use of other enhancement strategies), interactions with friends were generally more likely to elicit the use of impression management strategies than were interactions with strangers. Similarly, Gardner and Martinko (1988) found that participants in a study of school principals were more likely to use impression management strategies (e.g., other enhancement, apologies) when interacting with more familiar, as opposed to less familiar, individuals. Finally, in a study that examined the relation between familiarity and willingness to exert peer pressure (both antisocial and prosocial) among adolescents, McPhee (1996) found that participants were more likely to exert pressure on friends than on acquaintances. Thus, if differences among the age groups in triad familiarity affected our results in any measurable way, we believe that the adolescent participants, who completed the battery of measures with acquaintances (as opposed to self-selected friends), should have demonstrated group induced shifts toward risk taking that were no greater than those observed among the adults. But, overall, this was not the case.

Finally, it is conceivable that members of the three age groups differed in their prior experience with the subject matter of the risk-orientation measures. Thus, age differences in performance on the risky decision-making and risk-taking measures might be construed as an artifact of differences between the age groups in prior experience. However, any differences in experience with the content of the risky decision-making questionnaire were likely limited to the individual dilemmas. Thus, although a given dilemma might have been relatively more familiar to a particular age group, overall, the items were balanced in such a way that no one age group should have been more familiar with all five dilemmas than any other age group. With respect to the risk-taking measure, Chicken, it is likely that the youths and adults had more first hand experience with driving than the adolescents. However, Chicken is a video game played from a third-person perspective not a driving simulation experienced from a first-person perspective. Although adolescents may have limited experience with driving, they have ample experience with video and computer games Additionally, by adolescence, individuals have spent a great deal of time riding in cars and are surely familiar with the potential consequences of failing to follow traffic signals. Nonetheless, in order to ensure equal familiarity with the potential consequences of running a yellow light, all participants observed a demonstration round prior to playing the game in which they watched the animated car crash after running a yellow light. Thus, we believe that differences in first-hand driving experience had minimal impact on performance.

It is also important to note the few instances in which our hypotheses were not supported. Specifically, we failed to find significant age main effects or two-way Age $\times$ Condition interaction effects on the risk preference measure. In conceptualizing the study, we assumed that those who gave greater weight to the benefits versus the costs of risky decisions should be more likely to take risks. As noted earlier, there are a number of studies that have found strong correlations between cost–benefit consideration and risk taking (e.g., Goldberg & Fischoff, 2000; Horvath & Zuckerman, 1993; Thorton, Gibbons, & Gerrard, 2002). However, the samples in these studies were composed primarily of adults, and there is some evidence to suggest that measures of risk preference may not predict risk taking in the same way among adolescents as among adults. Indeed, despite findings that adolescents are more likely than are adults to engage in risky behavior, several studies suggest that adolescents are relatively similar to adults in their ability to recognize the risks and benefits of their actions. For example, Beyth-Marom et al. (1993) found few differences between adolescents and adults in the spontaneous mention of the costs and benefits associated with several risky actions. This has prompted some to argue that age differences in risky behavior may be better accounted for by differences in psychosocial functioning than by differences in more cognitive aspects of risk orientation, such as risk preference (Cauffman, 1996; Cauffman & Steinberg, 2000; Steinberg & Cauffman, 1996). In this respect, our failure to find age-related differences in individuals' cost–benefit appraisals is not entirely surprising.

We did find some interesting gender differences in risk preference, however. Specifically, males, particularly at younger ages, were more likely than were females to weigh the benefits of risky activities over the costs. Additionally, peer effects on benefit versus cost consideration were greater among males than among females. Although we did not explicitly predict these gender differences, our findings are consistent with several previous studies. For instance, Parsons, Halkitis, Bimbi, and Borkowski (2000) found that, among young adults, males reported more benefits and fewer risks when asked about the consequences of risky behaviors. Additionally, Brown et al. (1986) found that, at least among adolescents, males are more susceptible to peer influence than are females in antisocial or risky situations. Nonetheless, it is inter-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

esting that these gender-related differences in risk–benefit consideration did not translate into gender differences on the more direct measures of risk taking or risky decision making.

We also found differences in risk orientation as a function of ethnicity. First, we found differences between White and non-White participants in risk taking, risk preference, and risky decision making—particularly among adolescents (ethnic differences in risk orientation among adults were small to negligible). However, the direction of these ethnic group differences varied across measures. Whereas non-White adolescents demonstrated greater risk taking and risk preference than did White adolescents, White adolescents demonstrated greater risky decision making than did non-White adolescents. This is not entirely surprising given that prior studies have identified differences in the direction of ethnicity effects for different risk behaviors. For example, there is evidence to suggest that minority adolescents (particularly African Americans) are more likely than are White adolescents to engage in risky sexual behavior (e.g., Koniak-Griffin & Brecht, 1995; Neumark-Sztainer et al., 1996; Santelli, Lowry, Brener, & Robin, 2000) and to participate in delinquent activities (e.g., Blum et al., 2000; Hawkins, Laub, & Lauritsen, 1998; Piquero & Buka, 2002). However, there is also evidence to suggest that White adolescents may take more risks than may non-White adolescents when substance use is the behavior of interest. Specifically, a number of studies have found that White adolescents engage in more alcohol and tobacco use than adolescents from many non-White ethnic groups (Best et al., 2001; Blum et al., 2000; Brannock, Schandler, & Oncley, 1990; Douglas & Collins, 1997). Accordingly, researchers studying adolescent risk taking must exercise caution in asserting that ethnic group differences on particular risk measures reflect more general patterns of ethnic group differences in risk taking overall.

Second, we found that peer effects on risk orientation varied across ethnic groups. Specifically, we found that the effects of peer presence on risk preference were greater among non-White than among White participants. Though effect size estimates were not entirely consistent, inspection of mean scores suggests that the same pattern of ethnic group differences may also exist for risk taking. However, if it is, in fact, the case that non-White, relative to White, individuals are more susceptible to peer influence in risky situations, this ethnic group difference appears to be largely limited to adolescence. Although peer effects on risk taking and risk preference were greater among non-White than among White adolescents, non-White adults demonstrated levels of resistance to peer influence that were equal to or greater than those demonstrated by White adults. Though few studies have examined ethnic group differences in the development of resistance to peer influence, there is tentative evidence to support the finding that, relative to White adolescents, non-White adolescents may be more susceptible to the influence of others when in risky situations. For instance, Zimmerman, Sprecher, Langer, & Holloway (1995) found that African American and Hispanic adolescent females were slightly less confident in their ability to refuse unwanted sex than White adolescent females. However, to our knowledge, no prior research has directly examined the possibility that ethnic group differences in susceptibility to peer influence in risky situations may diminish as individuals move into adulthood. Thus, further research is necessary in order to both replicate and explain this finding.

In conclusion, it appears that differences in rates of group risk taking among adolescents versus adults are not simply the product of differences in the amount of time teenagers and adults spend with peers but are instead the result of age differences in individuals' orientation toward risky behavior when in the presence of friends. Moreover, our results suggest that the psychosocial capacities that undergird the ability to resist peer pressure may continue to develop throughout late adolescence and into early adulthood. Thus, interventions aimed at reducing risky behavior among adolescents and young adults—particularly those from ethnic minority groups—ought to focus some attention on increasing individuals' resistance to peer influence. For reasons not yet understood, the presence of peers makes adolescents and youth, but not adults, more likely to take risks and more likely to make risky decisions.

References

Andrews, J. A., Tildesley, E., Hops, H., & Li, F. (2002). The influence of peers on young adult substance use. *Health Psychology, 21,* 349–357.

Arnett, J. (1992). Reckless behavior in adolescence: A developmental perspective. *Developmental Review, 12,* 339–373.

Benthin, A., Slovic, P., & Severson, H. (1993). A psychometric study of adolescent risk perception. *Journal of Adolescence, 16,* 153–168.

Berndt, T. J. (1979). Developmental changes in conformity to peers and parents. *Developmental Psychology, 15,* 608–616.

Best, D., Rawaf, S., Rowley, J., Floyd, K., Manning, V., & Strang, J. (2001). Ethnic and gender differences in drinking and smoking among London adolescents. *Ethnicity & Health, 6,* 51–57.

Beyth-Marom, R., Austin, L., Fischoff, B., Palmgren, C., & Jacobs-Quadrel, M. (1993). Perceived consequences of risky behaviors: Adults and adolescents. *Developmental Psychology, 29,* 549–563.

Blascovich, J., & Ginsburg, G. P. (1974). Risky shifts and gambling: What's at stake? *Personality and Social Psychology Bulletin, 1,* 246–248.

Blascovich, J., Ginsburg, G. P., & Howe, R. C. (1975). Blackjack and the risky shift, II: Monetary stakes. *Journal of Experimental Social Psychology, 11,* 224–232.

Blascovich, J., Veach, T. L., & Ginsburg, G. P. (1973). Blackjack and the risky shift. *Sociometry, 36,* 42–55.

Blum, R. W., Beuhring, T., Shew, M. L., Bearinger, L. H., Sieving, R. E., & Resnick, M. D. (2000). The effects of race/ethnicity, income, and family structure on adolescent risk behaviors. *American Journal of Public Health, 90,* 1879–1884.

Bohra, K. A., & Pandey, J. (1984). Ingratiation toward strangers, friends, and bosses. *The Journal of Social Psychology, 122,* 217–222.

Brannock, J. C., Schandler, S. L., & Oncley, P. R. (1990). Cross-cultural and cognitive factors examined in groups of adolescent drinkers. *Journal of Drug Issues, 20,* 427–442.

Brown, B. (2004). Adolescents' relationships with peers. In R. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (2nd ed., pp. 363–394). New York: Wiley.

Brown, B. B., Clasen, D. R., & Eicher, S. A. (1986). Perceptions of peer pressure, peer conformity dispositions, and self-reported behavior among adolescents. *Developmental Psychology, 22,* 521–530.

Cauffman, E. (1996). Maturity of judgment: Psychosocial factors in adolescent decision making (Doctoral dissertation, Temple University, 1986). *Dissertation Abstracts International, 57(9-B),* 5943.

Cauffman, E., & Steinberg, L. (2000). Researching adolescents' judgment and culpability. In T. Grisso & R. G. Schwartz (Eds.), *Youth on trial: A developmental perspective on juvenile justice* (pp. 325–343). Chicago: The University of Chicago.

Cohen, S. L., & Ruis, C. B. (1974). Wary shift or risky shift? *Bulletin of the Psychonomic Society, 3,* 214–216.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Douglas, K. A., & Collins, J. L. (1997). Results from the 1995 National College Risk Behavior Survey. *Journal of American College Health, 46,* 55–66.

Erickson, M., & Jensen, G. F. (1977). Delinquency is still a group behavior! Toward revitalizing the group premise of the sociology of deviance. *Journal of Criminal Law and Criminology, 68,* 262–273.

Fischhoff, B. (1992). Risk taking: A developmental perspective. In J. F. Yates (Ed.), *Risk Taking Behavior* (pp. 133–162). Oxford, England: Wiley.

Ford, M., Wentzel, K., Wood, D., Stevens, E., & Siesfeld, G. A. (1990). Processes associated with integrative social competence: Emotional and contextual influences on adolescent social responsibility. *Journal of Adolescent Research, 4,* 405–425.

Fromme, K., Stroot, E., & Kaplan, D. (1993). The Comprehensive Effects of Alcohol questionnaire: Development and psychometric evaluation of a new expectancy questionnaire. *Psychological Assessment, 5,* 19–26.

Furby, L., & Beyth-Marom, R. (1992). Risk taking in adolescence: A decision-making perspective. *Developmental Review, 12,* 1–44.

Gardner, W. L., & Martinko, M. J. (1988). Impression management: An observational study linking audience characteristics with verbal self-presentations. *Academy of Management Journal, 31,* 42–65.

Goldberg, J., & Fischoff, B. (2000). The long-term risks in the short-term benefits: Perceptions of potentially addictive activities. *Health Psychology, 19,* 299–303.

Hawkins, D. F., Laub, J. H., & Lauritsen, J. L. (1998). Race, ethnicity, and serious juvenile offending. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pg. 30–46). Thousand Oaks, CA: Sage

Hensley, W. E. (1977). Probability, personality, age, and risk taking. *The Journal of Psychology, 95,* 139–145.

Hogg, M. A., Turner, J. C., & Davidson, B. (1990). Polarized norms and social frames of reference: A test of the self-categorization theory of group polarization. *Basic & Applied Social Psychology, 11,* 77–100.

Horvath, P., & Zuckerman, M. (1993). Sensation seeking, risk appraisal, and risky behavior. *Personality and Individual Differences, 14,* 41–52.

Kogan, N., & Wallach, M. A. (1967). Group risk taking as a function of members' anxiety and defensiveness levels. *Journal of Personality, 35,* 50–63.

Koniak-Griffin, D., & Brecht, M. (1995). Linkages between sexual risk taking, substance use, and AIDS knowledge among pregnant adolescents and young mothers. *Nursing Research, 44,* 340–346.

Lamm, H. (1967). Will an observer advise higher risk taking after hearing a discussion of the decision problem? *Journal of Personality and Social Psychology, 6,* 467–471.

Lamm, H., Trommsdorff, G., & Rost-Schaude, E. (1972). Self-image, perception of peers' risk acceptance and risky shift. *European Journal of Social Psychology, 2,* 255–272.

Lavery, B., Siegel, A. W., Cousins, J. H., & Rubovitz, D. S. (1993). Adolescent risk taking: An analysis of problem behavior in problem children. *Journal of Experimental Child Psychology, 55,* 277–294.

Leary, M. R., Nezlek, J. B., Downs, D., Radford-Davenport, J., Martin, J., & McMullen, A. (1994). Self-presentation in everyday interactions: Effects of target familiarity and gender composition. *Journal of Personality and Social Psychology, 67,* 664–673.

McBride, D. C., Weatherby, N. L., Inciardi, J. A., & Gillespie, S. A. (1999). AIDS susceptibility in a migrant population: Perception and behavior. *Substance Use and Misuse, 34,* 633–652.

McPhee, J. H. (1996). Influence strategies in young adolescent dyads. *Dissertation Abstracts International, 57*(02), 1468B.

Neumark-Sztainer, D., Story, M., French, S., Cassuto, N., Jacobs, D. R., & Resnick, M. D. (1996). Patterns of health-compromising behaviors among Minnesota adolescents: Sociodemographic variations. *American Journal of Public Health, 86,* 1599–1606.

Parsons, J. T., Halkitis, P. N., Bimbi, D., & Borkowski, T. (2000). Perceptions of the benefits and costs associated with condom use and unprotected sex among late adolescent college students. *Journal of Adolescence, 23,* 377–391.

Pilkonis, P. A., & Zanna, M. P. (1973). The choice-shift phenomenon in groups: Replication and extension. *Representative Research in Social Psychology, 4,* 36–47.

Piquero, A. R., & Buka, S. L. (2002). Linking juvenile and adult patterns of criminal activity in the Providence cohort of the National Collaborative Perinatal Project. *Journal of Criminal Justice, 30,* 259–272.

Pruitt, D. G., & Teger, A. I. (1969). The risky shift in group betting. *Journal of Experimental Social Psychology, 5,* 115–126.

Santelli, J. S., Lowry, R., Brener, N. D., & Robin, L. (2000). The association of sexual behaviors with socioeconomic status, family structure, and race/ethnicity among US adolescents. *American Journal of Public Health, 90,* 1582–1588.

Scott, E., Reppucci, N. D., & Woolard, J. (1995). Evaluating adolescent decisionmaking in legal contexts. *Law and Human Behavior, 19,* 221–244.

Sheldrick, R. C. (2004). Social networks and degree of psychopathy among adolescent offenders (Doctoral dissertation, Temple University, 1990). *Dissertation Abstracts International, 65(3-B),* 1564.

Singer, M., Dai, H., Weeks, M. R., & Malave, D. (1998). AIDS risk perception among women drug users in Hartford, CT. *Women, Drug Use, and HIV Infection, 27,* 67–85.

SPSS, Incorporated. (2005). Linear mixed-effects modeling in SPSS: An introduction to the MIXED procedure (LMEMWP-0305). Chicago, IL: Author.

Steinberg, L. (2004). Risk taking in adolescence: What changes, and why? *Annals of the NewYork Academy of Sciences, 1021,* 51–58.

Steinberg, L., & Cauffman, E. (1996). Maturity of judgment in adolescence: Psychosocial factors in adolescent decision making. *Law and Human Behavior, 20,* 249–272.

Steinberg, L., & Scott, E. (2003). Less guilty by reason of adolescence: Developmental immaturity, diminished responsibility, and the juvenile death penalty. *American Psychologist, 58,* 1009–1018.

Steinberg, L., & Silverberg, S. B. (1986). The vicissitudes of autonomy in early adolescence. *Child Development, 57,* 841–851.

Thorton, B., Gibbons, F. X., & Gerrard, M. (2002). Risk perception and prototype perception: Independent processes predicting risk behavior. *Personality and Social Psychology Bulletin, 28,* 986–999.

Vidmar, N. (1970). Group composition. *Journal of Experimental Social Psychology, 6,* 153–166.

Vinokur, A. (1971). Review and theoretical analysis of the effects of group processes upon individual and group decisions involving risk. *Psychological Bulletin, 76,* 231–250.

Wallach, M. A., & Kogan, N. (1965). The roles of information, discussion, and consensus in group risk taking. *Journal of Experimental Social Psychology, 1,* 1–19.

Yinon, Y., Jaffe, Y., & Feshbach, S. (1975). Risky aggression in individuals and groups. *Journal of Personality and Social Psychology, 31,* 808–815.

Zaleska, M. (1974). The effects of discussion on group and individual choices among bets. *European Journal of Social Psychology, 4,* 229–250.

Zimmerman, R. S., Sprecher, S., Langer, L. S., & Holloway, C. D. (1995). Adolescents' perceived ability to say "no" to unwanted sex. *Journal of Adolescent Research, 10,* 383–399.

Zimring, F. E. (1998). *American youth violence.* Oxford, England: Oxford University Press.

Received April 6, 2004
Revision received November 12, 2004
Accepted November 16, 2004 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Correction to Gardner and Steinberg (2005)

The article "Peer Influence on Risk Taking, Risk Preference, and Risky Decision Making in Adolescence and Adulthood: An Experimental Study," by Margo Gardner and Laurence Steinberg (*Developmental Psychology*, 2005, Vol. 41, No. 4, pp. 625–635), contains several typographical errors. On p. 630, in the first full paragraph of the first column, the degrees of freedom value that is reported as 293 should instead be 288. Additionally there are several typographical errors in the descriptive statistics table (Table 2). These errors do not have implications for the findings reported in the text of the article, nor do they alter the conclusions drawn from the study. The corrected table appears below.

Table 2
Descriptive Statistics for Group and Sole Participant Conditions by Age Group

		White			Non-White			Total		
Age	Condition	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*
					Risk taking (Chicken)					
Adolescent	Sole	−0.164	0.612	22	−0.035	0.722	29	−0.090	0.674	51
	Group	0.140	0.886	25	0.907	1.300	29	0.552	1.182	54
Youth	Sole	−0.258	0.729	24	−0.091	0.666	27	−0.170	0.694	51
	Group	0.139	0.848	30	0.289	0.893	20	0.199	0.860	50
Adult	Sole	−0.367	0.387	22	−0.316	0.620	19	−0.343	0.502	41
	Group	−0.080	0.566	24	−0.335	1.167	30	−0.221	0.949	54
					Risky decision making (YDMQ)					
Adolescent	Sole	2.127	0.511	22	1.862	0.571	29	1.977	0.557	51
	Group	2.072	0.519	25	2.021	0.522	29	2.044	0.516	54
Youth	Sole	2.142	0.564	24	2.074	0.448	27	2.106	0.502	51
	Group	2.506	0.477	32	2.282	0.358	22	2.415	0.443	54
Adult	Sole	1.655	0.339	22	1.926	0.758	19	1.781	0.581	41
	Group	1.625	0.397	24	1.720	0.497	30	1.678	0.454	54
					Risk preference (BRPM)					
Adolescent	Sole	1.391	0.335	22	1.372	0.287	29	1.380	0.305	51
	Group	1.312	0.383	25	1.835	0.921	29	1.593	0.764	54
Youth	Sole	1.517	0.521	24	1.393	0.370	27	1.451	0.447	51
	Group	1.819	0.600	32	1.555	0.365	22	1.711	0.529	54
Adult	Sole	1.427	0.317	22	1.368	0.354	19	1.400	0.332	41
	Group	1.425	0.355	24	1.393	0.358	30	1.407	0.354	54

Note. Chicken means are based on standardized scores. YDMQ = Youth Decision-Making Questionnaire; BRPM = Benthin Risk Perception Measure.

DOI: 10.1037/a0026993

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Molecular Psychiatry (2006) 11, 903–913
© 2006 Nature Publishing Group All rights reserved 1359-4184/06 $30.00
www.nature.com/mp

ORIGINAL ARTICLE

MAOA, maltreatment, and gene–environment interaction predicting children's mental health: new evidence and a meta-analysis

J Kim-Cohen[1,2], A Caspi[2,3], A Taylor[2], B Williams[2], R Newcombe[2], IW Craig[2] and TE Moffitt[2,3]

[1]*Department of Psychology, Yale University, New Haven, CT, USA;* [2]*Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, UK and* [3]*Department of Psychology, University of Wisconsin at Madison, Madison, WI, USA*

Previous research on adults has shown that a functional polymorphism in the promoter region of the monoamine oxidase A (*MAOA*) gene moderates the impact of childhood maltreatment on risk for developing antisocial behavior. Thus far, attempts to replicate this finding have been mixed. The current study (i) presents new data investigating this finding in a sample of 975 seven-year-old boys, and (ii) evaluates the extant data by conducting a meta-analysis of published findings. We replicated the original finding by showing that the *MAOA* polymorphism moderates the development of psychopathology after exposure to physical abuse, we extended the finding to childhood closer in time to the maltreatment experience, and we ruled-out the possibility of a spurious finding by accounting for passive and evocative gene–environment correlation. Moreover, meta-analysis demonstrated that across studies, the association between maltreatment and mental health problems is significantly stronger in the group of males with the genotype conferring low vs high *MAOA* activity. These findings provide the strongest evidence to date suggesting that the *MAOA* gene influences vulnerability to environmental stress, and that this biological process can be initiated early in life.
Molecular Psychiatry (2006) **11,** 903–913. doi:10.1038/sj.mp.4001851; published online 27 June 2006

Keywords: monoamine oxidase A; promoter polymorphism; maltreatment; psychopathology; antisocial behavior; ADHD; emotional problems

Introduction

Children who experience familial adversity such as physical abuse and inter-parental violence show wide variability in their mental health outcomes.[1,2] Many of these children develop behavioral and emotional difficulties, but many others are resilient and exhibit better functioning than predicted given their exposure to adversity. Recent findings[3] suggest that one explanation for variability in outcomes among maltreated individuals relates to a gene–environment interaction (G × E) involving a functional polymorphism in the promoter region of the monoamine oxidase A (*MAOA*) gene. Specifically, maltreated children with the *MAOA* genotype conferring low levels of the *MAOA* enzyme more often developed conduct disorder, antisocial personality, and violent criminality in adulthood than maltreated children with a high-activity *MAOA* genotype. *MAOA* selectively degrades serotonin, norephinephrine, and dopamine following reuptake from the synaptic cleft, and, therefore, plays a key role in regulating behavior.[4,5] Thus far, efforts to replicate Caspi *et al.*'s[3] results have met with mixed results[6–9] indicating that, as yet, accepting the *MAOA* by maltreatment G × E hypothesis would be premature without further investigation. The present study tested whether Caspi *et al.*[3] original finding would replicate in a representative birth cohort sample of 7-year-old boys.

Previous studies have examined the moderating influence of the *MAOA* polymorphism on psychopathology years after the exposure to adversity. Mental health outcomes as a consequence of childhood adversity have been measured at mean ages ranging from 13 to 26 years.[3,6,7] Whether the moderating effect of *MAOA* activity might be present in early life and closer in time to the stressful experience is unknown, yet this information is crucial for understanding the role that G × E might play in developmental processes. The present study asked whether *MAOA* gene activity moderates the impact of maltreatment on mental health outcomes in childhood.

A downward extension of the G × E hypothesis to young children requires a developmentally informed approach for assessing mental health.[10] Based on evidence from animal[5,11] and human[12,13] studies that linked *MAOA* enzyme deficiency with increased

Correspondence: Dr J Kim-Cohen, Department of Psychology, Yale University, PO Box 208205, New Haven, CT 06520, USA.
E-mail: julia.kim-cohen@yale.edu
Received 17 November 2005; revised 12 May 2006; accepted 17 May 2006; published online 27 June 2006

aggression, Caspi *et al*.'s[3] original G × E investigation of an adult sample targeted antisocial behavior as the hypothesized outcome. Subsequent studies[6–9] using adolescent samples did the same. However, longitudinal data have indicated that adolescent and adult violence and criminality have childhood origins often of the same kind (i.e., antisocial behavior), but also of a different type (e.g., anxiety, hyperactivity).[14] This is because early in development, psychopathology tends to be less differentiated than in later years, when broad behavioral and emotional difficulties develop into more stable and differentiated patterns of maladaptation.[15] Moreover, physical maltreatment is known to elevate risk for a variety of mental health outcomes, including but not limited to antisocial behavior.[16] In childhood, the most common domains of mental health problems include antisocial behavior such as aggression and destructiveness, attention deficits, impulsivity, and hyperactivity, and emotional problems such as anxiety and social withdrawal.[17] More often than not, these domains tend to co-occur in the same child[18] and when they do, this signals greater severity of psychopathology as well as a poorer long-term prognosis.[19] Accordingly, the combination of these mental health problems in 7-year-old boys was the focus of the present investigation.

In order to apply a rigorous test of a G × E, there must be evidence that the risk factor of interest and the outcome are linked via true environmental causation.[10] Otherwise, the association between a putative 'environmental' pathogen and a disorder could be mediated by an unknown third variable, which could reflect unidentified genetic influences. Such a correlation between genetic susceptibility and an environmental risk variable is referred to as a genotype–environment correlation (rGE).[20] The association between familial adversity and mental health outcomes might be explained by rGE of two different kinds.[21] First, aggressive parents may transmit to their children both an adverse rearing environment and a genetic susceptibility toward developing psychopathology (passive rGE). Second, a child with a particular genotype may behave in ways that elicit harsh treatment (evocative rGE). With respect to maltreatment and children's antisocial behavior, evidence of a causal association exists,[22] which argues against rGE as the only mediating mechanism. However, it is desirable to account for possible rGE and so far, only one G × E study involving *MAOA* activity and familial adversity has done so.[6] In the present study, we tested for the presence of passive rGE by controlling for mothers' antisocial personality and evocative rGE by testing whether boys' *MAOA* genotype predicted their exposure to maltreatment.

An extension of the replication strategy is to use meta-analysis systematically to aggregate results over multiple studies, which with regard to this particular G × E hypothesis remains mixed. One possible reason for discrepancy is across-study variability in conceptualizing and measuring the putative environmental risk factor. For instance, in a positive replication by Foley *et al*.,[6] 'family adversity' was measured prospectively as parental neglect, domestic violence, and harsh discipline. In the study by Haberstick *et al*.,[7] a partial failure to replicate, maltreatment was assessed retrospectively using different indicators. Meta-analysis is one tool for ascertaining whether a finding transcends such differences across studies. Meta-analysis is an increasingly valuable method in the field of psychiatric genetics, which in recent years has been plagued by non-replications,[23] leaving questions about gene effects on behavior in doubt. By pooling data from several studies, meta-analysis maximizes power to detect effects and avoids overemphasizing estimates from any single study.[24] The present investigation evaluated the current state of the cumulative evidence regarding *MAOA* activity, childhood adversity, and G × E via a meta-analysis.

Materials and methods

Sample

Participants are members of the Environmental Risk (E-Risk) Longitudinal Twin Study. The E-Risk sampling frame was two consecutive birth cohorts (1994 and 1995) in a birth register of twins born in England and Wales.[25] Of the 15 906 twin pairs born in these 2 years, 71% joined the register. Bias from non-response was corrected as follows.

The E-Risk Study probability sample was drawn using a high-risk stratification sampling procedure. High-risk families were those in which the mother had her first birth when she was 20 years of age or younger. We used this sampling (a) to replace high-risk families who were selectively lost to the register via non-response and (b) to ensure sufficient base rates of children growing up in at-risk environments. Age at first childbearing was used as the risk-stratification variable because it was recorded for virtually all families in the register, it is relatively free of measurement error, and early childbearing is a known risk factor for children's problem behaviors.[26,27] The sampling strategy resulted in a final sample in which one-third of Study mothers (younger only; $N = 314$) constitute a 160% oversample of mothers who were at high risk based on their young age at first birth (15–20 years). The other two-thirds of Study mothers ($N = 802$) accurately represent all mothers in the general population (aged 15–48) in England and Wales in 1994–1995 (estimates derived from the General Household Survey[28]). To provide unbiased statistical estimates that can be generalized to the population of British families with children born in the 1990s, the data reported in this article were corrected with weighting to represent the proportion of young mothers in that population.

The E-Risk Study sought a sample size of 1100 families to allow for attrition in future years of the longitudinal study while retaining statistical power. An initial list of families who had same-sex twins was drawn from the register to target for home visits. Of

the families from the initial list, 1116 (93%) participated in home-visit assessments when the twins were aged 5 years, forming the base sample for the study: 4% of families refused and 3% could not be reached after many attempts. Written informed consent was obtained from mothers. With parent's permission, questionnaires were posted to the children's teachers, and teachers returned questionnaires for 94% of cohort children.

A follow-up home visit was conducted 18–24 months after the children's age-5 assessment. Follow-up data were collected for 98% of the 1116 E-Risk Study families. At this follow-up, teacher questionnaires were obtained for 91% of the 2232 E-Risk Study children (93% of those taking part in the follow-up). The E-Risk Study has received ethical approval from the Maudsley Hospital Ethics Committee.

The present study includes 975 of the 1092 total boys in the E-Risk Study who are of Caucasian ancestry and for whom genotypic data were available. (DNA was not available for 36 boys. These boys did not differ significantly from the rest of the sample on risk for physical abuse exposure or level of mental health problems. Genotyping failure occurred in 23 cases.) Boys' single X chromosome yields two straightforwardly characterized *MAOA* promoter genotypes: high activity (66.3% in this sample) and low activity (33.7%).[3] Girls were excluded because, having two copies of the X chromosome, they fall into two homozygous groups, 'high–high' (39.1% in this sample) and 'low–low' (15.1%), and a third heterozygous group, 'high–low' (45.8%). Based on presently available evidence, girls' status on *MAOA* gene expression cannot be characterized with certainty because of conflicting findings regarding the inactivation status of the *MAOA* locus on the X chromosome.[29,30] In other words, whether the level of *MAOA* transcription in female subjects results from one or both copies of the *MAOA* gene is presently unknown and therefore, inferences about high or low *MAOA* activity in girls cannot be made.

DNA extraction and genotyping

At ages 5 and 7 years, DNA samples were obtained from study members via buccal swabs and extracted using a procedure described by Freeman *et al.*[31] Primer sequences are described by Sabol *et al.*,[4] namely *MAO APT1* (5′-ACAGCCTGACCGTGGA GAAG-3′) and *MAO APB1* (5′-GAACGGACGCTC CATTCGGA-3′). *MAO APT1* was 5′-labeled with the FAM fluorophore. Polymerase chain reaction (PCR) was carried out on a PTC-225 DNA engine (MJ Research, Hercules, CA, USA), using the following cycling conditions: initial 2 min denaturing step at 95°C, followed by 35 cycles of 94°C for 1 min, 55.5°C for 1 min and 72°C for 2 min, and a final extension phase of 72°C for 5 min. Reactions contained 1 × reaction Buffer IV (Abgene, Epsom, UK), 1.5 mM $MgCl_2$, 50 ng of genomic DNA, 5 pmols of each primer, 0.3 mM dNTPs, and 1.5 U of Native *Taq* (Promega, Madison, WI, USA) in a volume of 10 μl. PCR products were denatured in highly deionized formamide and analyzed by electrophoresis on an Applied Biosystems 3100 genetic analyzer (Applied Biosystems, Foster City, CA, USA), set up in genotyping mode, using a POP4 polymer and ROX-labeled GS500 size standard (Applied Biosystems). Results were analyzed using GeneScan v3.7 and Genotyper v3.6 software (Applied Biosystems).

Physical abuse exposure

Children's physical abuse exposure was assessed separately for each twin at the first and follow-up assessments by interviewing mothers with the standardized clinical interview protocol from the *Multi-Site Child Development Project*.[32] The interview protocol was designed to enhance mothers' comfort with reporting valid child maltreatment information, while also meeting researchers' legal and ethical responsibilities for reporting. Under the UK Children Act,[33] our responsibility was to secure intervention if maltreatment was current and ongoing. Such intervention on behalf of E-Risk families was carried out with parental cooperation in all but one case.

The protocol included standardized probe questions such as, 'Do you remember any time when (boy's name) was disciplined severely enough that he may have been hurt?' and 'Did you worry that you or someone else may have harmed or hurt (boy's name)?' Questions were carefully worded to avoid implying that the mother was the perpetrator, so mothers might feel more willing to report that a child had been maltreated. In cases where mothers reported any maltreatment, interviewers probed for details about the incident and recorded notes. Based on the mothers' narrative, interviewers coded if the boy had definitely been physically maltreated ($N = 34$). Examples of such maltreatment included being a victim of adjudicated assault by a teenaged sibling, punished by being burned with matches, had injuries (e.g., fractures or dislocations) from neglectful or abusive parental care, and/or were formally registered with a social services child protection team for physical abuse.

A further 147 boys were designated as 'probable maltreatment.' The 'probable maltreatment' category does not represent a milder form of abuse compared to the 'definite maltreatment' category. Instead, this group includes children who could only be suspected of having experienced maltreatment. In these probable cases, the boy had been reported by concerned schools, neighbors, and/or family members to child protective services but the case was not resolved or registered, he seemed afraid of his father during our home visit, the mother reported that he received frequent physical discipline, or she said he had been smacked harder than intended, leaving a mark or bruise. Some of these boys will not have experienced maltreatment, whereas a subset of them has been maltreated but our coding had to remain uncertain. To enhance certainty, we used information about

intimate partner violence to re-classify children in the probable category who were at greatest likelihood of having been maltreated. Researchers have documented the high correlation between intimate partner violence and child maltreatment.[34,35]

Intimate partner violence in the home was assessed by inquiring about 12 acts of physical violence (e.g., kicking a partner, threatening a partner with a knife) following the protocol of the *Conflict Tactics Scale – Form R*.[36] Mothers were asked about their own violence toward a partner and about any partner's violence toward them in the years of the child's life before the age-5 and the age-7 assessments. A methodological study demonstrated strong reliability and validity for the partner violence scale (i.e., 0.89 internal consistency and 75% inter-partner agreement).[37]

Children in the 'probable' maltreatment category were classified as having physical abuse exposure if their families were in the top 5% of the distribution of partner violence at either the age-5 or age-7 assessments. This method resulted in 28 boys from the 'probable maltreatment' group being added to the group of children we classified for this article as having been exposed to physical abuse. The resulting group totaled 62 boys (6.4% of the cohort unweighted; 4.7% weighted to represent the population) who were exposed to physical abuse.

Children's mental health outcomes

At the age-7 assessment, children's behavior problems were assessed with the *Child Behavior Checklist*[38] and the *Teacher Report Form*,[39] supplemented with items from the Rutter Child Scale[40] and additional items measuring *Diagnostic and Statistical Manual of Mental Disorders, 4th edition*[41] criteria for attention deficit hyperactivity disorder (ADHD). The CBCL cutoff for the clinical range (top 2% of the standardization sample) was exceeded by 7% of E-Risk cohort boys. Symptoms and behaviors were reported for the preceding 6 months and each item was scored as (0) 'not true,' (1) 'somewhat true,' and (2) 'very often true.' Sample items from the *Antisocial Behavior Scale* include 'physically attacks people,' 'lying or cheating,' and 'destroys things that belong to others' ($M = 21.57$, s.d. = 18.02). Sample items from the *Attention Deficit Hyperactivity Scale* include 'very restless,' 'cannot concentrate,' and 'impulsive or acts without thinking' ($M = 16.25$, s.d. = 12.40). Sample items from the *Emotional Problems Scale* include 'too fearful or anxious,' 'unhappy, sad, or depressed,' and 'withdrawn from social interaction' ($M = 11.08$, s.d. = 8.23). Following recommendations from the test manual,[42] we used mother interviews and teacher reports of children's behavior in combination to maximize reliability and validity. Mother and teacher reports of the same behaviors were moderately correlated and alpha reliabilities for all scales exceeded 0.85. Because simple combinatorial rules work as well, or better, than more complicated ones,[43] mothers' and teachers' reports were averaged to create scales of antisocial behavior, attention-deficit hyperactivity, and emotional problems. Principal Components Analysis (PCA) of the three subscales identified a single *Composite Mental Health Problem Scale* that accounted for 64.1% of the variance. Each of the component scales loaded adequately on the factor, with factor loadings ranging from 0.61 to 0.90. All scale scores were standardized with a mean of 0 and a standard deviation of 1.

Maternal antisocial personality symptoms

Mothers reported on their own lifetime antisocial history. Questions were derived from the *Diagnostic Interview Schedule*,[44] supplemented by items from the *Young Adult Behavior Checklist*.[45] The items covered illegal behavior, deceitfulness, impulsivity, aggressiveness, recklessness, and irresponsibility. A symptom was considered to be present if the mother reported behavioral items representing the symptom as being 'very true or often true.' Symptom counts in this study ranged from 0 to 6 ($M = 0.61$, s.d. = 1.07).[46]

Statistical analyses

First, we used ordinary least-squares regression to test the main effect associations of boys' physical abuse exposure and *MAOA* genotype in predicting mental health problems. Second, we tested for an interaction between *MAOA* activity and physical abuse exposure in predicting boys' mental health problems. Third, we tested whether boys' *MAOA* genotype was associated with risk for physical abuse exposure (i.e., evocative rGE). Fourth, we controlled for a possible passive rGE by adjusting for the effect of maternal antisocial personality symptoms in the regression analyses. Analyzing two boys in each family creates dependence in the data, and thus all regression results are based on the sandwich or Huber/White variance estimator,[47] a method available in *STATA 9.0*,[48] which adjusts estimated standard errors to account for that dependence and provides statistical tests that are robust to model assumptions.[49] Fifth, we conducted a meta-analysis pooling results from previous G × E studies of the *MAOA* polymorphism, physical maltreatment, and mental health with findings from the present study.

Results

Allele frequencies

Genotypes in the sample consisted of five variants of the 30-bp repeat sequence: 2- (0.2%), 3- (31.9%), 3.5- (2.1%), 4- (64.2%), and 5-repeats (1.6%). These allele frequencies matched closely the frequencies reported in other Caucasian samples.[31] In terms of expression, all studies agree on the functional classification of the two most common alleles, that is, 3-repeats (low activity), and 4-repeats (high activity). Of rare alleles, both Sabol *et al*.[4] and Deckert *et al*.[50] assayed the 3.5-repeat with the same result (high activity), whereas a discrepancy resulted for the 5-repeat. We chose the classification of Sabol *et al*.[4] (i.e., 5-repeat equals low

activity) as they assayed three cell lines as opposed to one. The rare 2-repeat, of which only two cases exist in our sample, was classified as low activity based upon precedence in previous studies. Dropping the 18 boys with the 2- or 5-repeat alleles did not alter the pattern of findings nor their significance.[3,6,7]

Predicting composite mental health problems

Sample sizes, group means, and effect size comparisons are presented in Table 1. As expected, there was a significant main effect of physical abuse exposure on children's mental health problems (Table 2). In addition, we found a significant main effect of *MAOA* activity ($P=0.017$), such that boys with the high activity allele had a higher level of mental health problems. The test for the interaction between *MAOA* activity and physical abuse exposure revealed a significant G × E ($b=-0.84$, s.e. = 0.40, $t=2.09$, $P=0.037$). This interaction showed that the effect of physical abuse exposure was significantly weaker among boys with high *MAOA* activity ($b=0.61$, s.e. = 0.22, $t=2.83$, $P=0.005$) than among boys with low *MAOA* activity ($b=1.45$, s.e. = 0.33, $t=4.40$, $P<0.001$) (see Figure 1).

Next, we examined whether there might be evidence of a rGE in two ways. First, we tested for the possibility that a child's genotype might be involved in evoking or eliciting physical abuse exposure. We found that it did not. Boys' *MAOA* activity was not significantly associated with likelihood of exposure to physical abuse (low-activity *MAOA* 3.5%; high-activity *MAOA* 5.3%; $F_{(1,974)}=1.89$, $P=0.170$). Second, we tested whether a child's *MAOA* genotype might be related to the likelihood of experiencing physical abuse indirectly via an association with

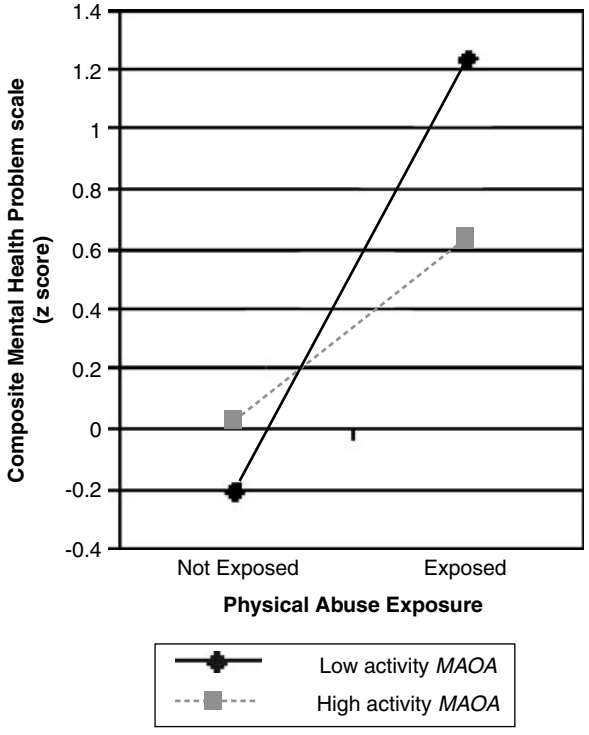


Figure 1 Gene-by-physical abuse exposure interaction predicting children's composite mental health problems.

Table 1 Sample sizes, means, standard deviations, and effect size (Cohen's *d*) comparisons of age-7 mental health outcomes in boys by *MAOA* genotype and physical abuse exposure

Children's mental health outcomes	*Physical abuse exposure*	*MAOA genotype*	
		Low activity	*High activity*
Sample sizes	Not exposed	$N=313$	$N=600$
	Exposed	$N=16$	$N=46$
		M (s.d.)	M (s.d.)
Composite mental health problem scale	Not exposed	−0.21 (0.91)	0.02 (0.97)
	Exposed	1.24 (1.38)	0.63 (1.04)
		$d=1.55$	$d=0.63$
Component mental health subscales			
Antisocial behavior subscale	Not exposed	−0.19 (0.90)	0.02 (0.98)
	Exposed	1.07 (1.61)	0.68 (1.02)
		$d=1.33$	$d=0.67$
Attentional problems and hyperactivity subscale	Not exposed	−0.17 (0.89)	0.04 (1.01)
	Exposed	1.02 (1.13)	0.45 (1.18)
		$d=1.32$	$d=0.40$
Emotional problem subscale	Not exposed	−0.13(0.90)	−0.03 (0.99)
	Exposed	0.92 (1.10)	0.37 (1.01)
		$d=1.15$	$d=0.40$

Abbreviations: MAOA, monoamine oxidase.

parental characteristics that increase environmental risk exposure. Maternal antisocial personality symptoms were significantly correlated with children's likelihood of physical abuse exposure ($b = 1.22$, s.e. = 0.26, $t = 4.79$, $P < 0.001$). However, when the model was adjusted for the main effects of physical abuse exposure, *MAOA* genotype, and maternal antisocial personality symptoms, the interaction between *MAOA* activity and physical abuse exposure remained significant (Table 2).

Do results hold for component mental health subscales?

We repeated the analyses separately predicting each of the component mental health subscales to examine whether the G × E interaction finding would be robust across multiple outcomes. All subscale findings of the interaction were in the predicted direction (Figure 2a–c) and mirrored the findings reported for the composite mental health measure. However, only the ADHD subscale yielded clear statistical significance ($b = -0.78$, s.e. = 0.35, $t = 2.26$, $P = 0.024$) (Table 2). This G × E finding remained significant ($P = 0.014$) after controlling for maternal antisocial personality symptoms.

Meta-analysis

We included studies in our meta-analysis if they fulfilled four criteria. First, the study had to be published in a peer-reviewed journal. Second, the study had to include genotypic information on the variable number tandem repeat polymorphism in the promoter region of the *MAOA* gene. Third, the study had to include a measure of serious familial adversity in childhood that was significantly associated in a main effect fashion with the outcome measure. Fourth, the sample had to be drawn from a non-clinical population. In addition to the present study, we know of four previous studies that meet all criteria.[3,6–8] A study by Young *et al*.[9] was excluded

Table 2 Results of regression analyses testing gene-by-physical abuse exposure interaction effects on children's mental health outcomes

Children's mental health outcomes	*Test for main effects*[a]				*Test for interaction term*				*Test for passive rGE*			
	b	*s.e.*	t	P	b	*s.e.*	t	P	b	*s.e.*	t	P
Composite mental health problem scale												
Constant	−0.19	0.07	2.83	0.005	−0.21	0.07	3.18	0.002	−0.31	0.07	4.56	0.001
Physical abuse exposure	0.83	0.19	4.44	0.001	1.45	0.33	4.41	0.001	1.25	0.30	4.12	0.001
MAOA genotype	0.19	0.08	2.40	0.017	0.23	0.08	2.81	0.005	0.22	0.08	2.86	0.004
Physical abuse exposure × *MAOA* genotype					−0.84	0.40	2.09	0.037	−0.85	0.39	2.19	0.029
Maternal antisocial personality symptoms									0.18	0.05	3.76	0.001
Component mental health subscales												
Antisocial behavior subscale												
Constant	−0.18	0.06	2.73	0.007	−0.19	0.06	3.01	0.003	−0.30	0.07	4.49	0.001
Physical abuse exposure	0.81	0.18	4.55	0.001	1.26	0.39	3.27	0.001	1.05	0.38	2.77	0.006
MAOA genotype	0.19	0.08	2.34	0.019	0.21	0.08	2.66	0.008	0.21	0.08	2.72	0.007
Physical abuse exposure × *MAOA* genotype					−0.61	0.44	1.38	0.169	−0.63	0.45	1.40	0.163
Maternal antisocial personality symptoms									0.18	0.05	3.84	0.001
ADHD subscale												
Constant	−0.15	0.06	2.42	0.016	−0.17	0.06	2.73	0.007	−0.26	0.06	4.07	0.001
Physical abuse exposure	0.62	0.18	3.37	0.001	1.20	0.27	4.49	0.001	1.01	0.23	4.35	0.001
MAOA genotype	0.18	0.08	2.23	0.026	0.21	0.08	2.57	0.010	0.21	0.08	2.62	0.009
Physical abuse exposure × *MAOA* genotype					−0.78	0.35	2.26	0.024	−0.80	0.32	2.48	0.014
Maternal antisocial personality symptoms									0.16	0.04	3.90	0.001
Emotional Problem Subscale												
Constant	−0.11	0.06	1.92	0.055	−0.13	0.06	2.20	0.028	−0.17	0.06	2.60	0.010
Physical abuse exposure	0.57	0.16	3.44	0.001	1.05	0.29	3.67	0.001	0.97	0.27	3.56	0.001
MAOA genotype	0.08	0.07	1.07	0.283	0.10	0.07	1.41	0.159	0.10	0.07	1.38	0.168
Physical abuse exposure × *MAOA* genotype					−0.65	0.35	1.85	0.065	−0.65	0.35	1.89	0.059
Maternal antisocial personality symptoms									0.06	0.04	1.65	0.099

[a]*N*s range between 954 and 959 across models because of missing data on boys' mental health measures or maternal antisocial personality symptoms.

Abbreviations: ADHD, attention deficit hyperactivity disorder; MAOA, monoamine oxidase A; rGE, gene–environment correlation.

The table presents final models with main effects and interactions entered simultaneously.

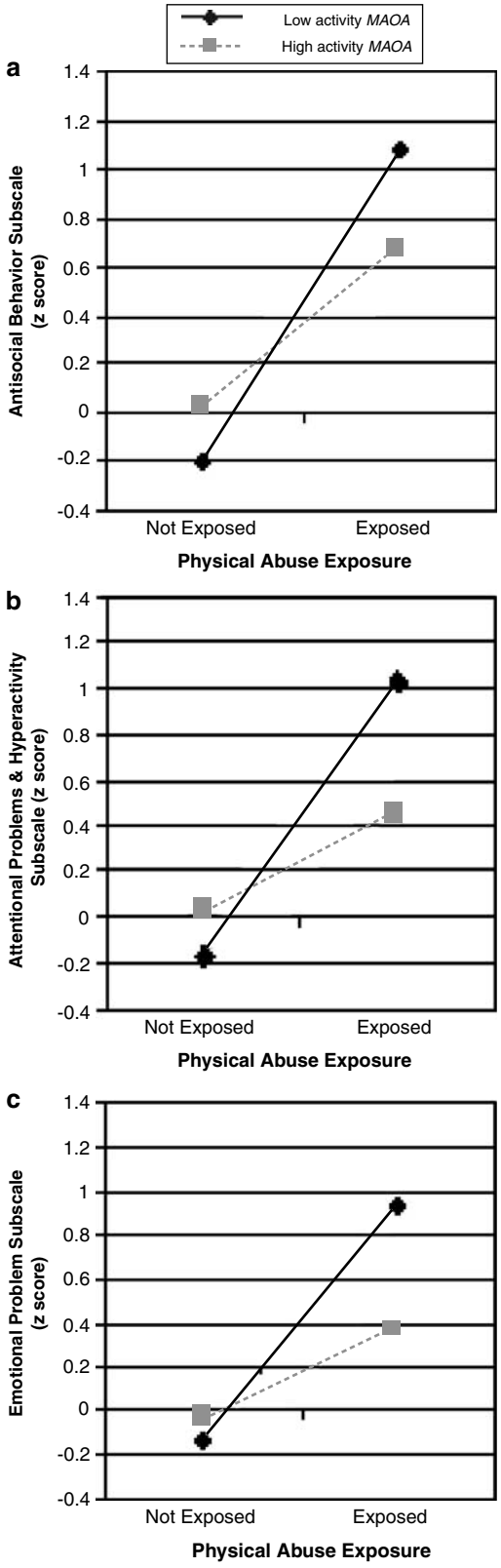


Figure 2 Gene-by-physical abuse exposure interaction predicting subscales of children's (**a**) antisocial behavior, (**b**) attentional problems and hyperactivity, and (**c**) emotional problems.

because their entire sample consisted of adolescents in clinical treatment for serious conduct problems, no matched control group was included, and their measure of lifetime conduct disorder symptoms places the temporal ordering of the risk and outcome variables in question.

Standard methods for effect size conversion were used to translate the results of the different studies into a common metric of correlations.[51,52] The correlations of interest were those reported between family adversity and antisocial behavior within each *MAOA* genotype group. If papers reported more than one outcome measure, we chose the most general measure of antisocial behavior (e.g., composite measures). Where necessary, authors were contacted for additional results in order to calculate effect sizes. All correlations were z-transformed and their standard errors derived.

Heterogeneity across studies was assessed using the χ^2-based Q statistic, as well as the I^2 measure of the percentage of across-study variation attributable to heterogeneity.[53,54] Using meta-analysis programs available in *STATA 9.0*,[48,55] fixed- and random-effects estimates and results were consistent across all analyses; we report only results from the more conservative random-effects model. Meta-analysis of the interaction effect was carried out using the difference in the transformed correlations across the *MAOA* groups within each study.

Figure 3 depicts results of the meta-analysis conducted in two ways. First, panel a presents a forest plot of the correlations between maltreatment and antisocial behavior by *MAOA* activity group for all studies. Within the high-activity *MAOA* group, there was no evidence of significant heterogeneity across studies ($\chi^2_{df=4}=0.79$, $P=0.940$, $I^2=0.0\%$), but there was mild–moderate heterogeneity across studies in the low-activity *MAOA* group as indicated by the I^2 measure ($\chi^2_{df=4}=6.29$, $P=0.179$, $I^2=36.4\%$). The random effects pooled estimates within *MAOA* activity groups were as follows: low *MAOA* activity, 0.32 (95% CI: 0.23, 0.40; $P<0.001$); high *MAOA* activity, 0.12 (95% CI: 0.08, 0.17; $P<0.001$).

Second, panel b presents a forest plot of the meta-analysis of the interaction effect based on the differences in correlations by *MAOA* activity group. No significant heterogeneity was detected across the studies ($\chi^2_{df=4}=3.79$, $P=0.440$, $I^2=0.0$). The pooled random-effects estimate of the change in correlations from the high- to low-activity *MAOA* groups indicated a significant effect of 0.18 (95% CI: 0.10, 0.26; $P<0.001$).

We carried out two additional sensitivity analyses on the meta-analysis results. First, we removed the study by Caspi *et al*.[3] and re-estimated the pooled effects to rule-out potential bias contributed by the first published study of this hypothesis.[24,56] This resulted in a significant but reduced pooled difference in correlations of 0.16 (95% CI: 0.07, 0.25; $P=0.001$). Second, we removed the Nilsson *et al*.[8] study, which reported an effect size larger in

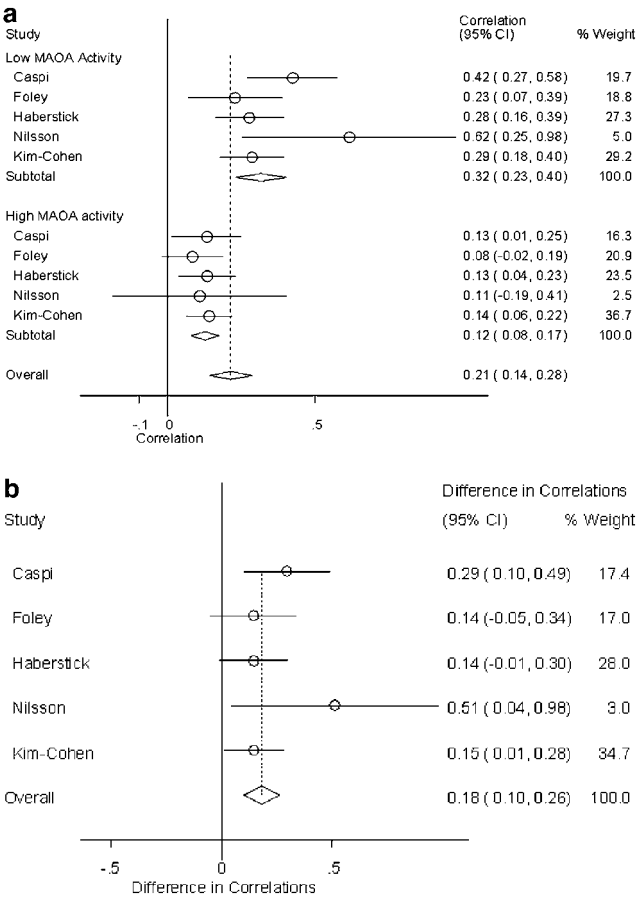


Figure 3 Meta-analysis pooling results across studies. Summary correlations between measures of childhood maltreatment and mental health are presented separately in the low- vs high-activity *MAOA* genotype groups. CI indicates confidence interval. Squares and % weight indicate the size of each study's contribution to the summary correlations indicated by diamonds. (**a**) Forest plot of correlations between childhood maltreatment and mental health, as a function of *MAOA* genotype group in five independent studies. (**b**) Forest plot of the interaction effect based on differences in correlations between childhood maltreatment and mental health observed in low- vs high-activity MAOA genotype groups in five independent studies.

magnitude than the original study by Caspi *et al*.[3] After both studies were removed, results showed a pooled effect of 0.15 (95% CI: 0.05, 0.24; $P=0.002$) (Meta-analysis using effect sizes for children's antisocial behavior from the Kim-Cohen *et al*. study resulted in the same pattern of significant findings. Details are available from first author.).

Discussion

Since 2002 when evidence first appeared that an *MAOA* polymorphism moderates the impact of childhood maltreatment on risk for antisocial behavior,[3] scientists have awaited supporting data through replication. The present investigation evaluated this G × E hypothesis in two ways. First, using data from a representative birth cohort sample of 7-year-old boys, we replicated and extended Caspi *et al*.'s[3] original results. Among children who were exposed to physical maltreatment, boys with the low-activity *MAOA* allele had mental health problem scores that were half a standard deviation higher than boys with the high-activity allele. Second, our preliminary meta-analysis found supportive evidence of this G × E effect. Pooling estimates from five studies, we found that the association between early familial adversity and mental health was significantly stronger in the low-activity *MAOA* vs the high-activity *MAOA* groups. If replication is the '*sine qua non* for accepting a hypothesis' (p. 627),[57] then evidence from the present study brings the field closer toward confirming that *MAOA* activity is meaningfully involved in explaining variability in developmental outcomes as a consequence of maltreatment.

Beyond replication and meta-analysis, this study contributes novel information in several additional ways. First, we extended the findings downward to a sample of 7-year-old boys, indicating that variability in young children's mental health is explained significantly by differences in vulnerability to stress as indexed by a genetic polymorphism. This finding suggests that one mechanism by which maltreatment leads to the development of psychopathology relates specifically to *MAOA* functioning, and that this biological process can be initiated early in life. Eventually, such evidence can inform not only what treatments might help prevent psychopathology in physically maltreated children but also when such intervention might be most successful. Second, we evaluated *MAOA* activity in relation to children's global mental health outcomes in addition to their antisocial behavior. In the Dunedin sample of adult males, the *MAOA* polymorphism did not moderate stress effects in predicting depression,[58] suggesting that in adults, *MAOA* activity moderates the impact of stress specifically towards an antisocial outcome. In childhood, the G × E involving *MAOA* appears to influence ADHD-related symptoms as well as a broader phenotype comprising several domains of mental health that together signal the beginning of a maladaptive trajectory toward the development of adult antisocial behavior. Third, we ruled-out the possibility that the G × E might be explained spuriously by a passive or evocative rGE.

One unexpected finding of the present study was that the *MAOA* polymorphism predicted children's mental health outcomes in a main effect fashion. Boys with the genotype conferring high *MAOA* activity had slightly but significantly elevated levels of global mental health problems as well as antisocial behavior and attention-deficits/hyperactivity relative to boys with the low-activity genotype. This finding is not inconsistent with previously published results of a significant main effect of the high-activity *MAOA* genotype on increasing risk for antisocial behavior and ADHD.[59–61] Moreover, closer inspection of other

G × E studies involving the *MAOA* polymorphism reveals similar findings, albeit at marginally significant levels.[3,6,7] For instance, Foley *et al*.[6] found a nonsignificant trend ($P < 0.14$) for the high-activity *MAOA* allele increasing risk for conduct disorder; after adjusting for the interaction between the *MAOA* polymorphism and maltreatment, the main effect became significant ($P = 0.04$). However, two recent studies found contradictory results with positive prediction from the low-activity *MAOA* genotype to ADHD.[62,63] Further research to resolve these discrepancies is needed to understand the direction of gene effects on conferring risk for psychopathology in the absence or presence of risk exposure.

Strengths of this study include a meta-analysis as well as an original analysis of data from a large, nonselected sample of boys at a young age when early emerging psychopathology predicts continuing difficulties in later life. Genotype, phenotype, and physical abuse exposure were measured using independent methods. The findings, however, should be interpreted in light of several limitations. First, our sample comprised Caucasian twins living in England and Wales. However, allele frequencies for the *MAOA* promoter polymorphism[3,4] and base rates of physical abuse exposure in our sample are comparable to those reported elsewhere.[34] Future research should test whether our findings replicate in singleton children and in non-Caucasian samples. Second, official records of maltreatment history were not available. However, unlike many large-scale studies that rely upon self-complete questionnaires, we interviewed caregivers face-to-face about violence exposure, a method having evidence of good validity.[22] Third, because girls' *MAOA* genotype cannot be characterized with confidence, our findings and the findings of previous studies are informative only about males. More progress toward understanding X-chromosome inactivation and *MAOA* gene expression is needed before the relevance of this finding to female subjects can be evaluated. Fourth, our meta-analysis contained only four studies in addition to Caspi *et al*.'s[3] original study. Therefore, our meta-analysis results should be re-evaluated once further tests of the hypothesis are published.

Ultimately, the goal of genomics research is to prevent mental disorder[57] and to refine treatment strategies.[64] Reliable G × E findings have considerable potential for informing such efforts.[10] However, a statistical interaction between a genotype and an environmental risk factor requires further research to uncover the biological mechanisms involved in the interaction.[65] Because the *MAOA* enzyme selectively metabolizes serotonin, norepinephrine, and dopamine,[4,5] which are involved in multiple brain functions associated with stress regulation,[66] it is likely to be one of myriad factors involved in the development of biological sensitivity to stress and the social context.[67] A statistical G × E involving *MAOA* thus represents an important launching pad for developmental neuroscience research into the underlying causal mechanisms involved in the etiology of psychopathology.[10,64,68–70] Moreover, it is possible that the *MAOA* gene may simply be a marker for a behavioral trait, which itself moderates the association between maltreatment and children's mental health.

Eradicating child maltreatment is clearly the preferred way to combat risk for psychiatric problems, and yet large numbers of children in Western societies are abused and exposed to family violence each year.[35] Once an adverse experience touches off an otherwise 'silent' genetic vulnerability and triggers a cascade of biological events toward atypical development, what can be done to halt or reverse the process? In addition to possible pharmacological interventions, recent research on the serotonin transporter polymorphism suggests that social support can protect even genetically vulnerable children from the negative sequelae of maltreatment.[71] As evidence for significant G × E in predicting mental health continues to emerge,[70] both scientists and the public are becoming increasingly aware that like many developmental processes,[72] the nature of gene effects on behavior, too, is often contingent upon experience.

Acknowledgments

We are grateful to the Study families and teachers for their participation. We thank Robert Plomin and Michael Rutter for their contributions, Thomas Achenbach for kind permission to adapt the CBCL, and members of the E-Risk team for their dedication, insights, and hard work. Terrie E Moffitt is a recipient of a Royal Society-Wolfson Research Merit Award. This research was supported by grants from the National Institute of Mental Health (MH45070 & MH49414), the UK Medical Research Council (G9806489 & G0100527), and the ESRC-SCOPIC Network.

References

1 Rutter M. The promotion of resilience in the face of adversity. In: Clarke-Stewart A, Dunn J (eds). *Families Count: Effects on Child and Adolescent Development*. Cambridge University Press: New York, 2006, pp 26–52.
2 Widom C, McGloin J. Resilience among abused and neglected children grown up. *Dev Psychopathol* 2001; **13**: 1021–1038.
3 Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Craig IW *et al*. Role of genotype in the cycle of violence in maltreated children. *Science* 2002; **297**: 851–854.
4 Sabol S, Hu S, Hamer D. A functional polymorphism in the monoamine oxidase A gene promoter. *Hum Genet* 1998; **103**: 273–279.
5 Shih J, Chen K, Ridd M. Monoamine oxidase: from genes to behavior. *Ann Rev Neurosci* 1999; **22**: 197–217.
6 Foley D, Eaves L, Wormley B, Silberg JL, Maes H, Kuhn J *et al*. Childhood adversity, monoamine oxidase A genotype, and risk for conduct disorder. *Arch Gen Psychiatry* 2004; **61**: 738–744.
7 Haberstick B, Lessem J, Hopfer C, Smolen A, Ehringer M, Timerlake D *et al*. Monoamine oxidase A (*MAOA*) and antisocial behaviors in the presence of childhood and adolescent maltreatment. *Am J Med Genet B Neuropsychiatr Genet* 2005; **135B**: 59–64.
8 Nilsson K, Sjoberg R, Damberg M, Leppert J, Ohrvik J, Alm P *et al*. Role of monoamine oxidase A genotype and psychosocial factors in male adolescent criminal activity. *Biol Psychiatry* 2005; **59**: 121–127.
9 Young S, Smolen A, Hewitt J, Haberstick M, Stallings M, Corley R *et al*. Interaction between *MAO-A* genotype and maltreatment in

risk for conduct disorder: failure to confirm in adolescent patients. *Am J Psychiatry* 2005; **163**: 1019–1025.
10 Moffitt TE, Caspi A, Rutter M. Strategy for investigating interactions between measured genes and measured environments. *Arch Gen Psychiatry* 2005; **62**: 473–481.
11 Cases O, Seif I, Grimsby J, Gaspar P, Chen K, Pournin S *et al.* Aggressive behavior and altered amounts of brain serotonin and norepinephrine in mice lacking *MAOA*. *Science* 1995; **268**: 1763–1766.
12 Brunner H, Nelen M, Breakefield X, Ropers H, van Oost B. Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. *Science* 1993; **262**: 578–580.
13 Brunner H, Nelen M, van Zandvoort P, Abeling N, van Gennip A, Wolters E *et al.* X-linked borderline mental retardation with prominent behavioral disturbance: phenotype, genetic localization, and evidence for disturbed monoamine metabolism. *Am J Hum Genet* 1993; **52**: 1032–1039.
14 Moffitt TE. Life course persistent versus adolescence-limited antisocial behavior. In: Cicchetti D, Cohen DJ (eds). *Developmental Psychopathology*, 2nd edn, vol. 3. Wiley: New York, 2006, pp 570–598.
15 Rutter M. Relationships between mental disorders in childhood and adulthood. *Acta Psychiatr Scand* 1995; **91**: 73–85.
16 Lansford JE, Dodge KA, Pettit GS, Bates JE, Crozier J, Kaplow J. A 12-year prospective study of the long-term effects of early child physical maltreatment on psychological, behavioral, and academic problems in adolescence. *Arch Pediatr Adolesc Med* 2002; **156**: 824–830.
17 Crijnen A, Achenbach T, Verhulst F. Problems reported by parents of children in multiple cultures: The Child Behavior Checklist syndrome constructs. *Am J Psychiatry* 1999; **156**: 569–574.
18 Caron C, Rutter M. Comorbidity in child psychopathology: concepts, issues and research strategies. *J Child Psychol Psychiatry* 1991; **32**: 1063–1080.
19 Verhulst FC, van der Ende J. 'Comorbidity' in an epidemiological sample: a longitudinal perspective.. *J Child Psychol Psychiatry* 1993; **34**: 767–783.
20 Rutter M, Silberg J. Gene–environment interplay in relation to emotional and behavioral disturbance. *Annu Rev Psychol* 2002; **53**: 463–490.
21 DiLalla L, Gottesman II. Biological and genetic contributors to violence – Widom's untold tale. *Psychol Bull* 1991; **109**: 125–129.
22 Jaffee SR, Moffitt TE, Caspi A, Taylor A. Physical maltreatment victim to antisocial chid: evidence of an environmentally mediated process. *J Abnorm Psychol* 2004; **113**: 44–55.
23 Insel TR, Collins FS. Psychiatry in the genomics era. *Am J Psychiatry* 2003; **160**: 616–620.
24 Ioannidis J, Ntzani E, Trikalinos T, Contopoulos-Ioannidis D. Replication validity of genetic association studies. *Nat Genet* 2001; **29**: 306–309.
25 Trouton A, Spinath FM, Plomin R. Twins Early Development Study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition, and behavior problems in childhood. *Twin Res* 2002; **5**: 444–448.
26 Maynard RA. *Kids Having Kids: Economic Costs and Social Consequences of Teen Pregnancy*. Urban Institute Press: Washington DC, 1997.
27 Moffitt TE, The E-Risk Study Team. Teen-aged mothers in contemporary Britain. *J Child Psychol Psychiatry* 2002; **43**: 727–742.
28 Bennett N, Jarvis L, Rowlands O, Singleton N, Haselden L. *Living in Britain: Results from the General Household Survey*. HMSO: London, 1996.
29 Benjamin D, Van Bakel I, Craig I. A novel expression based approach for assessing the inactivation status of human X-linked genes. *Eur J Hum Genet* 2000; **8**: 103–108.
30 Carrel L, Willard H. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 2005; **434**: 400–404.
31 Freeman B, Smith N, Curtis C, Huckett L, Mill J, Craig I. DNA from buccal swabs recruited by mail: evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping. *Behav Genet* 2003; **33**: 67–72.
32 Dodge KA, Pettit GS, Bates JE, Valente E. Social information-processing patterns partially mediate the effect of early physical abuse on later conduct problems. *J Abnorm Psychol* 1995; **104**: 632–643.
33 Department of Health. *The Children Act*. HMSO: London, 1989.
34 Osofsky J. Prevalence of children's exposure to domestic violence and child maltreatment: implications for prevention and intervention. *Clin Child Fam Psychol Rev* 2003; **6**: 161–170.
35 Tolan P, Gorman-Smith D, Henry D. Family violence. *Annu Rev Psychol* 2006; **57**: 557–583.
36 Straus MA. Measuring intrafamily conflict and violence: the Conflict Tactics (CT) Scales. In: Straus MA, Gelles RJ (eds). *Physical Violence in American Families: Risk Factors and Adaptations to Violence in 8145 Families*. Transaction: New Brunswick, NJ, 1990, pp 403–424.
37 Moffitt TE, Caspi A, Krueger RF, Magdol L, Margolin G, Silva PA *et al.* Do partners agree about abuse in their relationship? A psychometric evaluation of interpartner agreement. *Psychol Assess* 1997; **9**: 47–56.
38 Achenbach TM. *Manual for the Child Behavior Checklist/4–18 and 1991 Profile*. University of Vermont Department of Psychiatry: Burlington, VT, 1991.
39 Achenbach TM. *Manual for the Teacher's Report Form and 1991 Profile*. University of Vermont Department of Psychiatry: Burlington VT, 1991.
40 Sclare I. *The Child Psychology Portfolio*. NFER-Nelson Publishing Company: Windsor, Berkshire, 1997.
41 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. American Psychiatric Association: Washington, DC, 1994.
42 Achenbach TM, Rescorla L. *Manual for ASEBA School-Age Forms & Profiles*. University of Vermont, Research Center for Children, Youth, & Families: Burlington, VT, 2001.
43 Piacentini JC, Cohen P, Cohen J. Combining discrepant diagnostic information from multiple sources: are complex algorithms better than simple ones? *J Abnorm Child Psychol* 1992; **20**: 51–63.
44 Robins L, Cottler L, Bucholz K, Compton W. *Diagnostic Interview Schedule for DSM-IV*. Washington University School of Medicine: St Louis, MO, 1995.
45 Achenbach TM. *Manual for the Young Adult Self-Report and Young Adult Behavior Checklist*. University of Vermont Department of Psychiatry: Burlington, VT, 1997.
46 Kim-Cohen J, Caspi A, Rutter M, Polo Tomas M, Moffitt TE. The caregiving environments provided to children by depressed mothers with or without an antisocial history. *Am J Psychiatry* 2006; **163**: 1009–1018.
47 Gould W, Scribney W. *Maximum Likelihood Estimation with STATA*. Stata Press: College Station, TX, 1999.
48 StataCorp. *Stata Statistical Software: Release 9.0*. Stata Corporation: College Station, TX, 2005.
49 Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. *Annu Rev Pub Health* 2002; **23**: 151–169.
50 Deckert J, Catalano M, Syagailo Y, Okladnova O, DiBella D, Nothen M *et al.* Excess of high activity monoamine oxidase A gene promoter alleles in female patients with panic disorder. *Hum Mol Genet* 1999; **8**: 621–624.
51 Hunter JE, Schmidt FL. *Methods of Meta-analysis: Correcting Error and Bias in Research Findings*, 2nd edn, Sage: Newbury Park, CA, 2004.
52 Lipsey MW, Wilson DB. *Practical Meta-Analysis*. Sage: Thousand Oaks, CA, 2001.
53 Egger M, Smith GD, Altman DG. *Systematic Reviews in Health Care: Meta-analysis in Context*, 2nd edn. BMJ Publications: London, 2001.
54 Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statist Med* 2002; **21**: 1539–1558.
55 Sterne J, Bradburn M, Egger M. Meta-analysis in StataTM. In: Egger M, Smith GD, Altman DG (eds). *Systematic Reviews in Health Care: Meta-Analysis in Context*, 2nd edn, BMJ Publishing: London, 2001, pp 347–369.
56 Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG, Ioannidis JP. Establishment of genetic associations for complex diseases is

independent of early study findings. *Euro J Hum Genet* 2004; **12**: 762–769.
57 Merikangas KR, Risch N. Will the genomics revolution revolutionize psychiatry? *Am J Psychiatry* 2003; **160**: 625–635.
58 Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H *et al.* Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 2003; **301**: 386–389.
59 Beitchman J, Mik H, Ehtesham S, Douglas L, Kennedy J. *MAOA* and persistent, pervasive childhood aggression. *Mol Psychiatry* 2004; **9**: 546–547.
60 Manor I, Tyano S, Mel E, Eisenberg J, Bachner-Melman R, Kotler M *et al.* Family-based and association studies of monoamine oxidase A and attention deficit hyperactivity disorder (ADHD): preferential transmission of the long promoter-region repeat and its association with impaired performance on a continuous performance test (TOVA). *Mol Psychiatry* 2002; **7**: 626–632.
61 Manuck S, Flory J, Ferrell R, Mann J, Muldoon M. A regulatory polymorphism of the monoamine oxidase-A gene may be associated with variability in aggression, impulsivity, and central nervous system serotonergic responsivity. *Psychiatry Res* 2000; **95**: 9–23.
62 Lawson D, Turic D, Langley K, Pay H, Govan C, Norton N *et al.* Association analysis of monoamine oxidase A and attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 2003; **116B**: 84–89.
63 Domschke K, Sheehan K, Lowe N, Kirley A, Mullins C, O'Sullivan R *et al.* Association analysis of the monoamine oxidase A and B genes with attention deficit hyperactivity disorder (ADHD) in an Irish sample: preferential transmission of the MAO-A 941G allele to affected children. *Am J Med Genet B Neuropsychiatr Genet* 2005; **134B**: 110–114.
64 Kendler K. 'A gene for…': the nature of gene action in psychiatric genetics. *Am J Psychiatry* 2005; **162**: 1243–1252.
65 Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001; **358**: 1356–1360.
66 Charney DS. Psychobiological mechanisms of resilience and vulnerability: implications for successful adaptation to extreme stress. *Am J Psychiatry* 2004; **161**: 195–216.
67 Boyce W, Ellis B. Biological sensitivity to context: I. An evoluationary-developmental theory of the origins and functions of stress reactivity. *Dev Psychopathol* 2005; **17**: 271–301.
68 Cicchetti D, Blender JA. A multiple-levels-of-analysis approach to the study of developmental processes in maltreated children. *Proc Nat Acad Sci USA* 2004; **101**: 17325–17326.
69 Meyer-Lindenberg A, Buckholtz JW, Kolachana B, Hariri AR, Pezawas L, Blasi G *et al.* Neural mechanisms of genetic risk for impulsivity and violence in humans. *Proc Nat Acad Sci USA* 2006; **103**: 6269–6274.
70 Rutter M, Moffitt TE, Caspi A. Gene–environment interplay and psychopathology: multiple varieties but real effects. *J Child Psychol Psychiatry* 2006; **47**: 226–261.
71 Kaufman J, Yang B, Douglas-Palumberi H, Houshyar S, Lipschitz D, Krystal J *et al.* Social supports and serotonin transporter gene moderate depression in maltreated children. *Proc Nat Acad Sci USA* 2004; **101**: 17316–17321.
72 Pollak SD. Experience-dependent affective learning and risk for psychopathology in children. *Ann NY Acad Sci* 2003; **1008**: 102–111.

Physiol Rev 95: 853–951, 2015
Published June 24, 2015; doi:10.1152/physrev.00023.2014

NEURONAL REWARD AND DECISION SIGNALS: FROM THEORIES TO DATA

Wolfram Schultz

Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom



Schultz W. Neuronal Reward and Decision Signals: From Theories to Data. *Physiol Rev* 95: 853–951, 2015. Published June 24, 2015; doi:10.1152/physrev.00023.2014.—Rewards are crucial objects that induce learning, approach behavior, choices, and emotions. Whereas emotions are difficult to investigate in animals, the learning function is mediated by neuronal reward prediction error signals which implement basic constructs of reinforcement learning theory. These signals are found in dopamine neurons, which emit a global reward signal to striatum and frontal cortex, and in specific neurons in striatum, amygdala, and frontal cortex projecting to select neuronal populations. The approach and choice functions involve subjective value, which is objectively assessed by behavioral choices eliciting internal, subjective reward preferences. Utility is the formal mathematical characterization of subjective value and a prime decision variable in economic choice theory. It is coded as utility prediction error by phasic dopamine responses. Utility can incorporate various influences, including risk, delay, effort, and social interaction. Appropriate for formal decision mechanisms, rewards are coded as object value, action value, difference value, and chosen value by specific neurons. Although all reward, reinforcement, and decision variables are theoretical constructs, their neuronal signals constitute measurable physical implementations and as such confirm the validity of these concepts. The neuronal reward signals provide guidance for behavior while constraining the free will to act.

I.	INTRODUCTION	853
II.	REWARD FUNCTIONS	854
III.	LEARNING	862
IV.	APPROACH AND CHOICE	887

I. INTRODUCTION

Rewards are the most crucial objects for life. Their function is to make us eat, drink, and mate. Species with brains that allow them to get better rewards will win in evolution. This is what our brain does, acquire rewards, and do it in the best possible way. It may well be the reason why brains have evolved. Brains allow multicellular organisms to move about the world. By displacing themselves they can access more rewards than happen to come along by chance, thus enhancing their chance of survival and reproduction. However, movement alone does not get them any food or mating partners. It is necessary to identify stimuli, objects, events, situations, and activities that lead to the best nutrients and mating partners. Brains make individuals learn, select, approach, and consume the best rewards for survival and reproduction and thus make them succeed in evolutionary selection. To do so, the brain needs to identify the reward value of objects for survival and reproduction, and then direct the acquisition of these reward objects through learning, approach, choices, and positive emotions. Sensory discrimination and control of movements serve this prime role of the brain. For these functions, nature has endowed us with explicit neuronal reward signals that process all crucial aspects of reward functions.

Rewards are not defined by their physical properties but by the behavioral reactions they induce. Therefore, we need behavioral theories that provide concepts of reward functions. The theoretical concepts can be used for making testable hypotheses for experiments and for interpreting the results. Thus the field of reward and decision-making is not only hypothesis driven but also concept driven. The field of reward and decision-making benefits from well-developed theories of behavior as the study of sensory systems benefits from signal detection theory and the study of the motor system benefits from an understanding of mechanics. Reward theories are particularly important because of the absence of specific sensory receptors for reward, which would have provided basic physical definitions. Thus the theories help to overcome the limited explanatory power of physical reward parameters and emphasize the requirement for behavioral assessment of the reward parameters studied. These theories make disparate data consistent and coherent and thus help to avoid seemingly intuitive but paradoxical explanations.

Theories of reward function employ a few basic, fundamental variables such as subjective reward value derived from measurable behavior. This variable condenses all crucial factors of reward function and allows quantitative formal-

0031-9333/15 Copyright © 2015 the American Physiological Society

ization that characterizes and predicts a large variety of behavior. Importantly, this variable is hypothetical and does not exist in the external physical world. However, it is implemented in the brain in various neuronal reward signals, and thus does seem to have a physical basis. Although sophisticated forms of reward and decision processes are far more fascinating than arcane fundamental variables, their investigation may be crucial for understanding reward processing. Where would we be without the discovery of the esoteric electron by J. J. Thompson 1897 in the Cambridge Cavendish Laboratory? Without this discovery, the microprocessor and the whole internet would be impossible. Or, if we did not know about electromagnetic waves, we might assume a newsreader sitting inside the radio while sipping our morning coffee. This review is particularly concerned with fundamental reward variables, first concerning learning and then related to decision-making.

The reviewed work concerns primarily neurophysiological studies on single neurons in monkeys whose sophisticated behavioral repertoire allows well detailed, quantitative behavioral assessments while controlling confounds from sensory processing, movements, and attention. Thus I am approaching reward processing from the point of view of the tip of a microelectrode, one neuron at a time, thousands of them over the years, in rhesus' brains with more than two billion neurons. I apologize to the authors whose work I have not been able to cite in full, as there is a large number of recent studies on the subject and I am selecting these studies by their contribution to the concepts being treated here.

II. REWARD FUNCTIONS

A. Proximal Reward Functions Are Defined by Behavior

We have sensory receptors that react to environmental events. The retina captures electromagnetic waves in a limited range. Optical physics, physical chemistry, and biochemistry help us to understand how the waves enter the eye, how the photons affect the ion channels in the retinal photoreceptors, and how the ganglion cells transmit the visual message to the brain. Thus sensory receptors define the functions of the visual system by translating the energy from environmental events into action potentials and sending them to the brain. The same holds for touch, pain, hearing, smell, and taste. If there are no receptors for particular environmental energies, we do not sense them. Humans do not feel magnetic fields, although some fish do. Thus physics and chemistry are a great help for defining and investigating the functions of sensory systems.

Rewards have none of that. Take rewarding stimuli and objects: we see them, feel them, taste them, smell them, or hear them. They affect our body through all sensory systems, but there is not a specific receptor that would capture the particular motivational properties of rewards. As reward functions cannot be explained by object properties alone, physics and chemistry are only of limited help, and we cannot investigate reward processing by looking at the properties of reward receptors. Instead, rewards are defined by the particular behavioral reactions they induce. Thus, to understand reward function, we need to study behavior. Behavior becomes the key tool for investigating reward function, just as a radio telescope is a key tool for astronomy.

The word reward has almost mystical connotations and is the subject of many philosophical treatises, from the ethics of the utilitarian philosophy of Jeremy Bentham (whose embalmed body is displayed in University College London) and John Stuart Mill to the contemporary philosophy of science of Tim Schroeder (39, 363, 514). More commonly, the man on the street views reward as a bonus for exceptional performance, like chocolate for a child getting good school marks, or as something that makes us happy. These descriptions are neither complete nor practical for scientific investigations. The field has settled on a number of well-defined reward functions that have allowed an amazing advance in knowledge on reward processing and have extended these investigations into economic decision-making. We are dealing with three, closely interwoven, functions of reward, namely, learning, approach behavior and decision-making, and pleasure.

1. Learning

Rewards have the potential to produce learning. Learning is Pavlov's main reward function (423). His dog salivates to a bell when a sausage often follows, but it does not salivate just when a bell rings without consequences. The animal's reaction to the initially neutral bell has changed because of the sausage. Now the bell predicts the sausage. No own action is required, as the sausage comes for free, and the learning happens also for free. Thus Pavlovian learning (classical conditioning) occurs automatically, without the subject's own active participation, other than being awake and mildly attentive. Then there is Thorndike's cat that runs around the cage and, among other things, presses a lever and suddenly gets some food (589). The food is great, and the cat presses again, and again, with increasing enthusiasm. The cat comes back for more. This is instrumental or operant learning. It requires an own action; otherwise, no reward will come and no learning will occur. Requiring an action is a major difference from Pavlovian learning. Thus operant learning is about actions, whereas Pavlovian learning is about stimuli. The two learning mechanisms can be distinguished schematically but occur frequently together and constitute the building blocks for behavioral reactions to rewards.

Rewards in operant conditioning are positive reinforcers. They increase and maintain the frequency and strength of the behavior that leads to them. The more reward Thorndike's cat gets, the more it will press the lever. Reinforcers do nt only strengthen and maintain behavior for the cat but also for obtaining stimuli, objects, events, activities, and situations as different as beer, whisky, alcohol, relaxation, beauty, mating, babies, social company, and hundreds of others. Operant behavior gives a good definition for rewards. Anything that makes an individual come back for more is a positive reinforcer and therefore a reward. Although it provides a good definition, positive reinforcement is only one of several reward functions.

2. Approach behavior and decision-making

Rewards are attractive. They are motivating and make us exert an effort. We want rewards; we do not usually remain neutral when we encounter them. Rewards induce approach behavior, also called appetitive or preparatory behavior, and consummatory behavior. We want to get closer when we encounter them, and we prepare to get them. We cannot get the meal, or a mating partner, if we do not approach them. Rewards usually do not come alone, and we often can choose between different rewards. We find some rewards more attractive than others and select the best reward. Thus we value rewards and then decide between them to get the best value. Then we consume them. So, rewards are attractive and elicit approach behavior that helps to consume the reward. Thus any stimulus, object, event, activity, or situation that has the potential to make us approach and consume it is by definition a reward.

3. Positive emotions

Rewards have the potential to elicit positive emotions. The foremost emotion evoked by rewards is pleasure. We enjoy having a good meal, watching an interesting movie, or meeting a lovely person. Pleasure constitutes a transient response that may lead to the longer lasting state of happiness. There are different degrees and forms of pleasure. Water is pleasant for a thirsty person, and food for a hungry one. The rewarding effects of taste are based on the pleasure it evokes. Winning in a big lottery is even more pleasant. But many enjoyments differ by more than a few degrees. The feeling of high that is experienced by sports people during running or swimming, the lust evoked by encountering a ready mating partner, a sexual orgasm, the euphoria reported by drug users, and the parental affection to babies constitute different forms (qualities) rather than degrees of pleasure (quantities).

Once we have experienced the pleasure from a reward, we may form a desire to obtain it again. When I am thirsty or hungry and know that water or food helps, I desire them. Different from such specific desire, there are also desires for imagined or even impossible rewards, such as flying to Mars, in which cases desires become wishes (514). Desire requires a prediction, or at least a representation, of reward and constitutes an active process that is intentional [in being about something (529)]. Desire makes behavior purposeful and directs it towards identifiable goals. Thus desire is the emotion that helps to actively direct behavior towards known rewards, whereas pleasure is the passive experience that derives from a received or anticipated reward. Desire has multiple relations to pleasure; it may be pleasant in itself (I feel a pleasant desire), and it may lead to pleasure (I desire to obtain a pleasant object). Thus pleasure and desire have distinctive characteristics but are closely intertwined. They constitute the most important positive emotions induced by rewards. They prioritize our conscious processing and thus constitute important components of behavioral control. These emotions are also called liking (for pleasure) and wanting (for desire) in addiction research (471) and strongly support the learning and approach generating functions of reward.

Despite their immense power in reward function, pleasure and desire are very difficult to assess in an objectively measurable manner, which is an even greater problem for scientific investigations on animals, despite attempts to anthropomorphize (44). We do not know exactly what other humans feel and desire, and we know even less what animals feel. We can infer pleasure from behavioral responses that are associated with verbal reports about pleasure in humans. We could measure blood pressure, heart rate, skin resistance, or pupil diameter as manifestations of pleasure or desire, but they occur with many different emotions and thus are unspecific. Some of the stimuli and events that are pleasurable in humans may not even evoke pleasure in animals but act instead through innate mechanisms. We simply do not know. Nevertheless, the invention of pleasure and desire by evolution had the huge advantage of allowing a large number of stimuli, objects, events, situations, and activities to be attractive. This mechanism importantly supports the primary reward functions in obtaining essential substances and mating partners.

4. Potential

Rewards have the potential to produce learning, approach, decisions, and positive emotions. They are rewards even if their functions are not evoked at a given moment. For example, operant learning occurs only if the subject makes the operant response, but the reward remains a reward even if the subject does not make the operant response and the reward cannot exert its learning function. Similarly an object that has the potential to induce approach or make me happy or desire it is a reward, without necessarily doing it every time because I am busy or have other reasons not to engage. Pavlovian conditioning of approach behavior, which occurs every time a reward is encountered as long as it evokes at least minimal attention, nicely shows this.

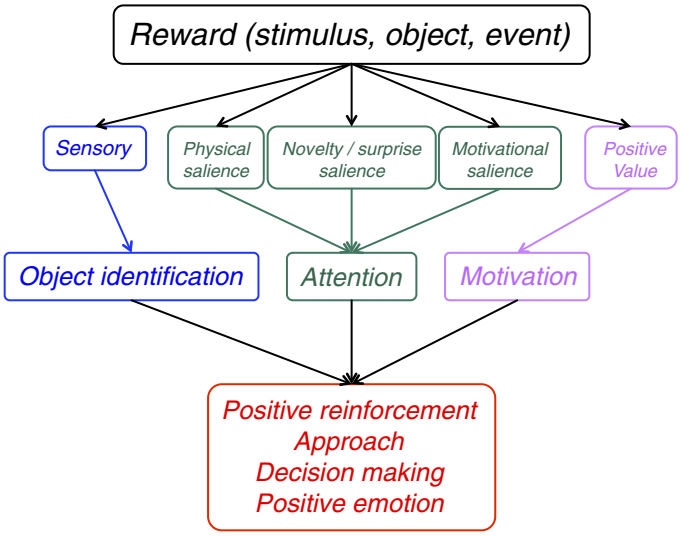


FIGURE 1. Reward components and their functions. The sensory component reflects the impact of environmental stimuli, objects, and events on the organism (blue). Pleasurable activities and situations belong also in this sensory component. The three salience components eliciting attentional responses (green) derive from the physical impact (*left*), novelty (*middle*), and commonly from reward and punishment (*right*). The specific positively motivating function of rewards derives from the value component (pink). Value does not primarily reflect physical parameters but the brain's subjective assessment of the usefulness of rewards for survival and reproduction. These reward components are either external (sensory, physical salience) or internal (generated by the brain; value, novelty/surprise salience, motivational salience). All five components together ensure adequate reward function.

5. Punishment

The second large category of motivating events besides rewards is punishment. Punishment produces negative Pavlovian learning and negative operant reinforcement, passive and active avoidance behavior and negative emotions like fear, disgust, sadness, and anger (143). Finer distinctions separate punishment (reduction of response strength, passive avoidance) from negative reinforcement (enhancing response strength, active avoidance).

6. Reward components

Rewarding stimuli, objects, events, situations, and activities consist of several major components. First, rewards have basic sensory components (visual, auditory, somatosensory, gustatory, and olfactory) (**FIGURE 1,** *left*), with physical parameters such as size, form, color, position, viscosity, acidity, and others. Food and liquid rewards contain chemical substances necessary for survival such as carbohydrates, proteins, fats, minerals, and vitamins, which contain physically measurable quantities of molecules. These sensory components act via specific sensory receptors on the brain. Some rewards consist of situations, which are detected by cognitive processes, or activities involving motor processes, which also constitute basic components analogous to sensory ones. Second, rewards are salient and thus elicit attention, which are manifested as orienting responses (**FIGURE 1,** *middle*). The salience of rewards derives from three principal factors, namely, their physical intensity and impact (physical salience), their novelty and surprise (novelty/surprise salience), and their general motivational impact shared with punishers (motivational salience). A separate form not included in this scheme, incentive salience, primarily addresses dopamine function in addiction and refers only to approach behavior (as opposed to learning) and thus to reward and not punishment (471). The term is at odds with current results on the role of dopamine in learning (see below) and reflects an earlier assumption of attentional dopamine function based on an initial phasic response component before distinct dopamine response components were recognized (see below). Third, rewards have a value component that determines the positively motivating effects of rewards and is not contained in, nor explained by, the sensory and attentional components (**FIGURE 1,** *right*). This component reflects behavioral preferences and thus is subjective and only partially determined by physical parameters. Only this component constitutes what we understand as a reward. It mediates the specific behavioral reinforcing, approach generating, and emotional effects of rewards that are crucial for the organism's survival and reproduction, whereas all other components are only supportive of these functions.

The major reward components together ensure maximal reward acquisition. Without the sensory component, reward discrimination would be difficult; without the attentional components, reward processing would be insufficiently prioritized; and without valuation, useless objects would be pursued. In practical reward experiments, the value component should be recognized as a distinct variable in the design and distinguished and uncorrelated from the sensory and attentional components.

The reward components can be divided into external components that reflect the impact of environmental stimuli, objects and events on the organism, and internal components generated by brain function. The sensory components are external, as they derive from external events and allow stimulus identification before evaluation can begin. In analogy, the external physical salience components lead to stimulus-driven attention. The foremost internal component is reward value. It is not inherently attached to stimuli, objects, events, situations, and activities but reflects the brain's assessment of their usefulness for survival and reproduction. Value cannot be properly defined by physical reward parameters but is represented in subjective preferences that are internal, private, unobservable, and incomparable between individuals. These preferences are elicited by approach behavior and choices that can be objectively measured. The internal nature of value extends to its associated motivational salience. Likewise, reward predictors are not

hardwired to outside events but require neuronal learning and memory processes, as does novelty/surprise salience which relies on comparisons with memorized events. Reward predictors generate top-down, cognitive attention that establishes a saliency map of the environment before the reward occurs. Further internal reward components are cognitive processes that identify potentially rewarding environmental situations, and motor processes mediating intrinsically rewarding movements.

B. Distal Reward Function Is Evolutionary Fitness

Modern biological theory conjectures that the currently existing organisms are the result of evolutionary competition. Advancing the idea about survival of the fittest organisms, Richard Dawkins stresses gene survival and propagation as the basic mechanism of life (114). Only genes that lead to the fittest phenotype will make it. The phenotype is selected based on behavior that maximizes gene propagation. To do so, the phenotype must survive and generate offspring, and be better at it than its competitors. Thus the ultimate, distal function of rewards is to increase evolutionary fitness by ensuring survival of the organism and reproduction. Then the behavioral reward functions of the present organisms are the result of evolutionary selection of phenotypes that maximize gene propagation. Learning, approach, economic decisions, and positive emotions are the proximal functions through which phenotypes obtain the necessary nutrients for survival, mating, and care for offspring.

Behavioral reward functions have evolved to help individuals to propagate their genes. Individuals need to live well and long enough to reproduce. They do so by ingesting the substances that make their bodies function properly. The substances are contained in solid and liquid forms, called foods and drinks. For this reason, foods and drinks are rewards. Additional rewards, including those used for economic exchanges, ensure sufficient food and drink supply. Mating and gene propagation is supported by powerful sexual attraction. Additional properties, like body form, enhance the chance to mate and nourish and defend offspring and are therefore rewards. Care for offspring until they can reproduce themselves helps gene propagation and is rewarding; otherwise, mating is useless. As any small edge will ultimately result in evolutionary advantage (112), additional reward mechanisms like novelty seeking and exploration widen the spectrum of available rewards and thus enhance the chance for survival, reproduction, and ultimate gene propagation. These functions may help us to obtain the benefits of distant rewards that are determined by our own interests and not immediately available in the environment. Thus the distal reward function in gene propagation and evolutionary fitness defines the proximal reward functions that we see in everyday behavior. That is why foods, drinks, mates, and offspring are rewarding.

The requirement for reward seeking has led to the evolution of genes that define brain structure and function. This is what the brain is made for: detecting, seeking, and learning about rewards in the environment by moving around, identifying stimuli, valuing them, and acquiring them through decisions and actions. The brain was not made for enjoying a great meal; it was made for getting the best food for survival, and one of the ways to do that is to make sure that people are attentive and appreciate what they are eating.

C. Types of Rewards

The term *reward* has many names. Psychologists call it positive reinforcer because it strengthens behaviors that lead to reward, or they call it outcome of behavior or goal of action. Economists call it a good or commodity and assess the subjective value for the decision maker as utility. We now like to identify the kinds of stimulus, object, event, activity, and situation that elicit the proximal functions of learning, approach, decision-making, and positive emotions and thus serve the ultimate, distal reward function of evolutionary fitness.

1. Primary homeostatic and reproductive rewards

To ensure gene propagation, the primary rewards mediate the survival of the individual gene carrier and her reproduction. These rewards are foods and liquids that contain the substances necessary for individual survival, and the activities necessary to mate, produce offspring, and care for the offspring. They are attractive and the main means to achieve evolutionary fitness in all animals and humans. Primary food and liquid rewards serve to correct homeostatic imbalances. They are the basis for Hull's drive reduction theory (242) that, however, would not apply to rewards that are not defined by homeostasis. Sexual behavior follows hormonal imbalances, at least in men, but is also strongly based on pleasure. To acquire and follow these primary alimentary and mating rewards is the main reason why the brain's reward system has evolved in the first place. Note that "primary" reward does not refer to the distinction between unconditioned versus conditioned reward; indeed, most primary rewards are learned and thus conditioned (foods are primary rewards that are typically learnt).

2. Nonprimary rewards

All other rewards serve to enhance the function of primary alimentary and mating rewards and thus enhance the chance for survival, reproduction, and evolutionary selection. Even though they are not homeostatic or reproductive rewards, they are rewards in their own rights. These nonprimary rewards can be physical, tangible objects like money, sleek cars, or expensive jewelry, or material liquids like a glass of wine, or particular ingredients like spices or alcohol. They can have particular pleasant sensory properties

like the visual features of a Japanese garden or a gorgeous sunset, the acoustic beauty of Keith Jarrett's Cologne Concert, the warm feeling of water in the Caribbean, the gorgeous taste of a gourmet dinner, or the irresistible odor of a perfume. Although we need sensory receptors to detect these rewards, their motivating or pleasing properties require further appreciation beyond the processing of sensory components **(FIGURE 1)**. A good example is Canaletto's Grand Canal **(FIGURE 2)** whose particular beauty is based on physical geometric properties, like the off-center Golden Ratio position reflecting a Fibonacci sequence (320). However, there is nothing intrinsically rewarding in this ratio of physical proportions. Its esthetic (and monetary) value is entirely determined by the subjective value assigned by our brain following the sensory processing and identification of asymmetry. Although we process great taste or smell as sensory events, we appreciate them as motivating and pleasing due to our subjective valuation. This rewarding function, cultivated in gourmet eating, enhances the appreciation and discrimination of high-quality and energy-rich primary foods and liquids and thus ultimately leads to better identification of higher quality food and thus higher survival chances (as gourmets are usually not lacking food, this may be an instinctive trait for evolutionary fitness). Sexual attraction is often associated with romantic love that, in contrast to straightforward sex, is not required for reproduction and therefore does not have primary reward functions. However, love induces attachment and facilitates care for offspring and thus supports gene propagation. Sexual rewards constitute also the most straightforward form of social rewards. Other social rewards include friendship, altruism, general social encounters, and societal activities that promote group coherence, cooperation, and competition which are mutually beneficial for group members and thus evolutionarily advantageous.

Nonphysical, nonmaterial rewards, such as novelty, gambling, jokes, suspense, poems, or relaxation, are attractive but less tangible than primary rewards. These rewards have no homeostatic basis and no nutrient value, and often do not promote reproduction directly. We may find the novelty of a country, the content of a joke, or the sequence of words in a poem more rewarding than the straightforward physical aspects of the country or the number of words in the joke or poem. But novelty seeking, and to some extent gambling, may help to encounter new food sources. Jokes, suspense, poems, and relaxation may induce changes of viewpoints and thus help to understand the world, which may help us to consider alternative food sources and mating partners, which is helpful when old sources dry up. Although these rewards act indirectly, they increase evolutionary fitness by enhancing the functions of primary alimentary and reproductive rewards.

Rewards can also be intrinsic to behavior (31, 546, 547). They contrast with extrinsic rewards that provide motivation for behavior and constitute the essence of operant behavior in laboratory tests. Intrinsic rewards are activities that are pleasurable on their own and are undertaken for their own sake, without being the means for getting extrinsic rewards. We may even generate our own rewards through internal decisions. Mice in the wild enter wheels and run on them on repeated occasions without receiving any other reward or benefit, like the proverbial wheel running hamster (358). Movements produce proprioceptive stimulation in muscle spindles and joint receptors, touch stimulation on the body surface, and visual stimulation from seeing the movement, all of which can be perceived as pleasurable and thus have reward functions. Intrinsic rewards are genuine rewards in their own right, as they induce learning, approach, and pleasure, like perfectioning, playing, and enjoying the piano. Although they can serve to condition higher order rewards, they are not conditioned, higher order rewards, as attaining their reward properties does not require pairing with an unconditioned reward. Other examples for intrinsic rewards are exploration, own beauty, gourmet eating, visiting art exhibitions, reading books, taking power and control of people, and investigating the natural order of the world. The pursuit of intrinsic rewards seems private to the individual but may inadver-



FIGURE 2. Subjective esthetic reward value derived from objective physical properties. The beauty of the Canaletto picture depends on the Golden Ratio of horizontal proportions, defined as $(a + b)/a = a/b \sim 0.618$; a and b for width of image. The importance of geometric asymmetry becomes evident when covering the left part of the image until the distant end of the canal becomes the center of the image: this increases image symmetry and visibly reduces beauty. However, there is no intrinsic reason why physical asymmetry would induce subjective value: the beauty appears only in the eye of the beholder. (Canaletto: The Upper Reaches of the Grand Canal in Venice, 1738; National Gallery, London.)

tently lead to primary and extrinsic rewards, like an explorer finding more food sources by venturing farther afield, a beauty queen instinctively promoting attractiveness of better gene carriers, a gourmet improving food quality through hightened culinary awareness, an artist or art collector stimulating the cognitive and emotional capacities of the population, a scholar providing public knowledge from teaching, a politician organizing beneficial cooperation, and a scientist generating medical treatment through research, all of which enhance the chance of survival and reproduction and are thus evolutionary beneficial. The double helix identified by Watson and Crick for purely scientific reasons is now beneficial for developing medications. The added advantage of intrinsic over solely extrinsic rewards is their lack of narrow focus on tangible results, which helps to develop a larger spectrum of skills that can be used for solving wider ranges of problems. Formal mathematical modeling confirms that systems incorporating intrinsic rewards outperform systems relying only on extrinsic rewards (546). Whereas extrinsic rewards such as food and liquids are immediately beneficial, intrinsic rewards are more likely to contribute to fitness only later. The fact that they have survived evolutionary selection suggests that their later benefits outweigh their immediate costs.

D. What Makes Rewards Rewarding?

Why do particular stimuli, objects, events, situations, and activities serve as rewards to produce learning, approach behavior, choices, and positive emotions? There are four separate functions and mechanisms that make rewards rewarding. However, these functions and mechanisms serve the common proximal and distal reward functions of survival and gene propagation. Individuals try to maximize one mechanism only to the extent that the other mechanisms are not compromised, suggesting that the functions and mechanisms are not separate but interdependent.

1. Homeostasis

The first and primary reward function derives from the need of the body to have particular substances for building its structure and maintaining its function. The concentration of these substances and their derivatives is finely regulated and results in homeostatic balance. Deviation from specific set points of this balance requires replenishment of the lost substances, which are contained in foods and liquids. The existence of hunger and thirst sensations demonstrates that individuals associate the absence of necessary substances with foods and liquids. We obviously know implicitly which environmental objects contain the necessary substances. When the blood sodium concentration exceeds its set point, we drink water, but depletion of sodium leads to ingestion of salt (472).

Two brain systems serve to maintain homeostasis. The hypothalamic feeding and drinking centers together with intestinal hormones deal with immediate homeostatic imbalances by rapidly regulating food and liquid intake (24, 46). In contrast, the reward centers mediate reinforcement for learning and provide advance information for economic decisions and thus are able to elicit behaviors for obtaining the necessary substances well before homeostatic imbalances and challenges arise. This preemptive function is evolutionarily beneficial as food and liquid may not always be available when an imbalance arises. Homeostatic imbalances are the likely source of hunger and thirst drives whose reduction is considered a prime factor for eating and drinking in drive reduction theories (242). They engage the hypothalamus for immediate alleviation of the imbalances and the reward systems for preventing them. The distinction in psychology between drive reduction for maintaining homeostasis and reward incentives for learning and pursuit may grossly correspond to the separation of neuronal control centers for homeostasis and reward. The neuroscientific knowledge about distinct hypothalamic and reward systems provides important information for psychological theories about homeostasis and reward.

The need for maintaining homeostatic balance explains the functions of primary rewards and constitutes the evolutionary origin of brain systems that value stimuli, objects, events, situations, and activities as rewards and mediate the learning, approach, and pleasure effects of food and liquid rewards. The function of all nonprimary rewards is built onto the original function related to homeostasis, even when it comes to the highest rewards.

2. Reproduction

In addition to acquiring substances, the other main primary reward function is to ensure gene propagation through sexual reproduction, which requires attraction to mating partners. Sexual activity depends partly on hormones, as shown by the increase of sexual drive with abstinence in human males. Many animals copulate only when their hormones put them in heat. Castration reduces sexual responses, and this deficit is alleviated by testosterone administration in male rats (146). Thus, as with feeding behavior, hormones support the reward functions involved in reproduction.

3. Pleasure

Pleasure is not only one of the three main reward functions but also provides a definition of reward. As homeostasis explains the functions of only a limited number of rewards, the prevailing reason why particular stimuli, objects, events, situations, and activities are rewarding may be pleasure. This applies first of all to sex (who would engage in the ridiculous gymnastics of reproductive activity if it were not for the pleasure) and to the primary homeostatic rewards of food and liquid, and extends to money, taste, beauty, social encounters and nonmaterial, internally set, and intrinsic

rewards. Pleasure as the main effect of rewards drives the prime reward functions of learning, approach behavior, and decision making and provides the basis for hedonic theories of reward function. We are attracted by most rewards, and exert excruciating efforts to obtain them, simply because they are enjoyable.

Pleasure is a passive reaction that derives from the experience or prediction of reward and may lead to a longer lasting state of happiness. Pleasure as hallmark of reward is sufficient for defining a reward, but it may not be necessary. A reward may generate positive learning and approach behavior simply because it contains substances that are essential for body function. When we are hungry we may eat bad and unpleasant meals. A monkey who receives hundreds of small drops of water every morning in the laboratory is unlikely to feel a rush of pleasure every time it gets the 0.1 ml. Nevertheless, with these precautions in mind, we may define any stimulus, object, event, activity, or situation that has the potential to produce pleasure as a reward.

Pleasure may add to the attraction provided by the nutrient value of rewards and make the object an even stronger reward, which is important for acquiring homeostatically important rewards. Once a homeostatic reward is experienced, pleasure may explain even better the attraction of rewards than homeostatis. Sensory stimuli are another good example. Although we employ arbitrary, motivationally neutral stimuli in the laboratory for conditioning, some stimuli are simply rewarding because they are pleasant to experience. The esthetic shape, color, texture, viscosity, taste, or smell of many rewards are pleasant and provide own reward value independently of the nutrients they contain (although innate and even conditioned mechanisms may also play a role, see below). Examples are changing visual images, movies, and sexual pictures for which monkeys are willing to exert effort and forego liquid reward (53, 124), and the ever increasing prices of paintings (fame and pride may contribute to their reward value). Not surprisingly, the first animal studies eliciting approach behavior by electrical brain stimulation interpreted their findings as discovery of the brain's pleasure centers (398), which were later partly associated with midbrain dopamine neurons (103, 155) despite the notorious difficulties of identifying emotions in animals.

4. Innate mechanisms

Innate mechanisms may explain the attractions of several types of reward in addition to homeostasis, hormones, and pleasure. A powerful example is parental affection that derives from instinctive attraction. Ensuring the survival of offspring is essential for gene propagation but involves efforts that are neither driven by homeostasis nor pleasure. As cute as babies are, repeatedly being woken up at night is not pleasurable. Generational attraction may work also in the other direction. Babies look more at human faces than at scrambled pictures of similar sensory intensity (610), which might be evolutionary beneficial. It focuses the baby's attention on particularly important stimuli, initially those coming from parents. Other examples are the sensory aspects of rewards that do not evoke pleasure, are nonnutritional, and are not conditioned but are nevertheless attractive. These may include the shapes, colors, textures, viscosities, tastes, or smells of many rewards (539), although some of them may turn out to be conditioned reinforcers upon closer inspection (640).

5. Punisher avoidance

The cessation of pain is often described as pleasurable. Successful passive or active avoidance of painful events can be rewarding. The termination or avoidance might be viewed as restoring a "homeostasis" of well being, but it is unrelated to proper vegetative homeostasis. Nor is avoidance genuine pleasure, as it is built on an adverse event or situation. The opponent process theory of motivation conceptualizes the reward function of avoidance (552), suggesting that avoidance may be a reward in its own right. Accordingly, the simple timing of a conditioned stimulus relative to an aversive event can turn punishment into reward (584).

E. Rewards Require Brain Function

1. Rewards require brains

Although organisms need sensory receptors to detect rewards, the impact on sensory receptors alone does not explain the effects of rewards on behavior. Nutrients, mating partners, and offspring are not attractive by themselves. Only the brain makes them so. The brain generates subjective preferences that reflect on specific environmental stimuli, objects, events, situations, and activities as rewards. These preferences are elicited by choices and quantifiable from behavioral reactions, typically choices but also reaction times and other measures. Reward function is explained by assuming the notion of value attributed to individual rewards. Value is not a physical property but determined by brain activity that interprets the potential effect of a reward on survival and reproduction. Thus rewards are internal to the brain and based entirely on brain function (547).

2. Explicit neuronal reward signals

Information processing systems work with signals. In brains, the signals that propagate through the circuits are the action potentials generated by each neuron. The output of the system is the observable behavior. In between are neurons and synapses that transmit and alter the signals. Each neuron works with thousands of messenger molecules and membrane channels that determine the action potentials. The number of action potentials, and somewhat their

pattern, varies monotonically with sensory stimulation, discrimination, and movements (3, 385). Thus the key substrates for the brain's function in reward are specific neuronal signals that occur in a limited number of brain structures, including midbrain dopamine neurons, striatum, amygdala, and orbitofrontal cortex **(FIGURE 3)**. Reward signals are also found in most component structures of the basal ganglia and the cerebral cortical areas, often in association with sensory or motor activity. The signals can be measured as action potentials by neurophysiology and are also reflected in transmitter concentrations assessed by electrochemistry (638) and as synaptic potentials detected by magnetic resonance imaging in blood oxygen level dependent (BOLD) signals (328). Whereas lesions in humans and animals demonstrate necessary involvements of specific brain structures in behavioral processes, they do not inform about the way the brain processes the information underlying these processes. Electric and optogenetic stimulation evokes action potentials and thus helps to dissect the influence of individual brain structures on behavior, but it does not replicate the natural signals that occur simultaneously in several interacting structures. Thus the investigation of neuronal signals is an important method for understanding crucial physiological mechanisms for the survival and reproduction of biological organisms.

3. Reward retina

Neuronal signals in sensory systems originate in specific receptors that define the signal content. However, rewards have no dedicated receptors. Neuronal processing would benefit from an explicit signal that identifies a reward irrespective of sensory properties and irrespective of actions required to obtain it. The signal might be analogous to visual responses of photoreceptors in the retina that constitute the first processing stage for visual perception. To obtain an explicit reward signal, the brain would extract the rewarding component from heterogeneous, polysensory environmental objects and events. A signal detecting the reward properties of an apple should not be concerned with its color unless color informs about reward properties of the fruit. Nor should it code the movement required to obtain the apple, other than assessing the involved effort as economic cost. External visual, somatic, auditory, olfactory, and gustatory stimuli predicting original, unconditioned rewards become conditioned rewards through Pavlovian conditioning. The issue for the brain is then to extract the reward information from the heterogeneous responses to the original and conditioned rewards and generate a common reward signal. Neurons carrying such a signal would constitute the first stage in the brain at which the reward property of environmental objects and events would be coded and conveyed to centers engaged in learning, approach, choice, and pleasure. Such abstract reward neurons would be analogous to the retinal photoreceptors as first visual processing stage (519).

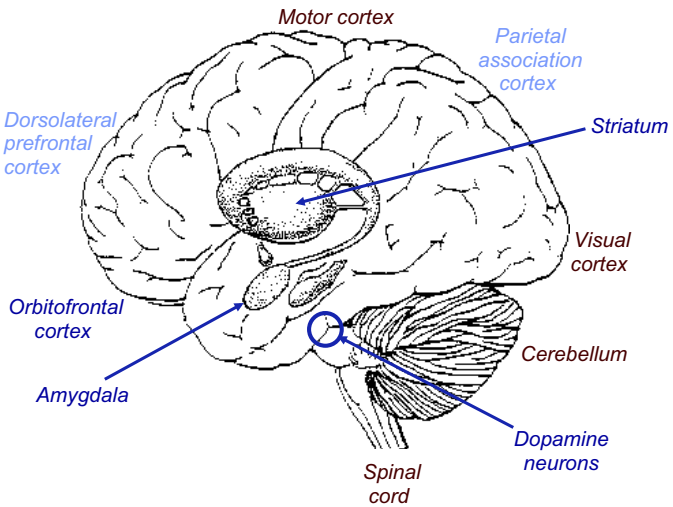


FIGURE 3. Principal brain structures for reward and decision-making. Dark blue: main structures containing various neuronal subpopulations coding reward without sensory stimulus or motor action parameters ("explicit reward signals"). Light blue: structures coding reward in conjunction with sensory stimulus or motor action parameters. Maroon: non-reward structures. Other brain structures with explicit or conjoint reward signals are omitted for clarity.

Despite the absence of specific reward receptors, there are chemical, thermal, and mechanical receptors in the brain, gut, and liver that detect important and characteristic reward ingredients and components, such as glucose, fatty acids, aromatic amino acids, osmolality, oxygen, carbon dioxide, temperature, and intestinal volume, filling, and contractions. In addition to these exteroceptors, hormone receptors are stimulated by feeding and sex (24, 46). These receptors are closest to being reward receptors but nevertheless detect only physical, sensory reward aspects, whereas reward value is still determined by internal brain activity.

The absence of dedicated receptors that by themselves signal reward value may not reflect evolutionary immaturity, as rewards are as old as multicellular organisms. Rather valuation separate from physical receptors may be an efficient way of coping with the great variety of objects that can serve as rewards at one moment or another, and of adapting reward value to changing requirements, including deprivation and satiation. Rather than having a complex, omnipotent, polysensory receptor that is sensitive to all possible primary and conditioned rewards and levels of deprivation, however infrequent they may occur, it might be easier to have a neuronal mechanism that extracts the reward information from the existing sensory receptors. The resulting neuronal reward signal would be able to detect rewarding properties in a maximum number of environmental objects, increase the harvest of even rare rewards, and relate their value to current body states, which all together enhance the chance of survival. Such a signal would be an efficient solution to the existential problem of benefitting from the

largest possible variety of rewards with reasonable hardware and energy cost.

III. LEARNING

A. Principles of Reward Learning

1. Advantage of learning

Learning is crucial for evolutionary fitness. It allows biological organisms to obtain a large variety of rewards in a wide range of environments without the burden of maintaining hard-wired mechanisms for every likely and unlikely situation. Organisms that can learn and adapt to their environments can live in more widely varying situations and thus acquire more foods and mating partners. Without learning, behavior for these many situations would need to be preprogrammed which would require larger brains with more energy demands. Thus learning saves brain size and energy and thus enhances evolutionary fitness. These advantages likely prompted the evolutionary selection of the learning function of rewards.

Learning processes lead to the selection of those behaviors that result in reward. A stimulus learned by Pavlovian conditioning elicits existing behavior when the stimulus is followed by reward. Natural behavioral reactions such as salivation or approach that improve reward acquisition become more frequent when a stimulus is followed by a reward. Male fish receiving a Pavlovian conditioned stimulus before a female approaches produce more offspring than unconditioned animals (223), thus demonstrating the evolutionary benefit of conditioning. Operant learning enhances the frequency of existing behavior when this results in reward. Thorndike's cat ran around randomly until it came upon a lever that opened a door to a food source. Then its behavior focused increasingly on the lever for the food. Thus Pavlovian and operant learning commonly lead to selection of behavior that is beneficial for survival.

Learning is instrumental for selecting the most beneficial behaviors that result in the best nutrients and mating partners in the competition for individual and gene survival. In this sense selection through learning is analogous to the evolutionary selection of the fittest genes. For both, the common principle is selection of the most efficient characteristics. The difference is on the order of time and scale. Selection of behavior through learning is based on outcome over minutes and hours, whereas selection of traits through evolution is based on survival of the individual. Evolutionary selection includes the susceptibility to reward and the related learning mechanisms that result in the most efficient acquisition of nutrients and mating partners.

2. Pavlovian learning

In Pavlov's experiment, the dog's salivation following the bell suggests that it anticipates the sausage. A Pavlovian conditioned visual stimulus would have a similar effect. The substances that individuals need to live are packaged in objects or liquids or are contained in animals they eat. We need to recognize a pineapple to get its juice that contains necessary substances like water, sugar, and fibers. Recognition of these packages could be hard wired into brain function, which would require a good number of neurons to detect a reasonable range of rewards. Alternatively, a flexible mechanism could dynamically condition stimuli and events with the necessary rewards and thus involve much less neurons representing the rewards. That mechanism makes individuals learn new packages when the environment changes. It also allows humans to manufacture new packages that were never encountered during evolution. Through Pavlovian conditioning humans learn the labels for hamburgers, baby food, and alcoholic beverages. Thus Pavlovian conditioning touches the essence of reward functions in behavior and allows individuals to detect a wide range of rewards from a large variety of stimuli while preventing run away neuron numbers and brain size. It is the simplest form of learning that increases evolutionary fitness and was thus selected by evolution.

Pavlovian conditioning makes an important conceptual point about reward function. We do not need to act to undergo Pavlovian conditioning. It happens without our own doing. But our behavior reveals that we have learned something, whether we wanted to or not. Pavlov's bell by itself would not make the dog salivate. But the dog salivates when it hears the bell that reliably precedes the sausage. From now on, it will always salivate to a bell, in particular when the bell occurs in the laboratory in which it has received all the sausages. Salivation is an automatic component of appetitive, vegetative approach behavior.

Although initially defined to elicit vegetative reactions, Pavlovian conditioning applies also to other behavioral reactions. The bell announcing the sausage elicits also approach behavior involving eye, limb, and licking movements. Thus Pavlovian learning concerns not only vegetative reactions but also skeletal movements. Furthermore, Pavlovian conditioning occurs in a large variety of behavioral situations. When operant learning increases actions that lead to reward, the more reward inadvertently conditions the involved stimuli in a Pavlovian manner. Thus Pavlovian and operant processes often go together in learning. When we move to get a reward, we also react to the Pavlovian conditioned, reward-predicting stimuli. In choices between different rewards, Pavlovian conditioned stimuli provide crucial information about the rewarded options. Thus Pavlovian and operant behavior constitute two closely linked forms of learning that extend well beyond laboratory experiments. Finally, Pavlovian conditioning concerns not

only the motivating components of rewards but also their attentional aspects. It confers motivational salience to arbitrary stimuli that elicits stimulus-driven attention and directs top-down attention to the reward and thus focuses behavior on pursuing and acquiring the reward. Taken together, Pavlovian conditioning constitutes a fundamental mechanism that is crucial for a large range of learning processes.

3. Reward prediction and information

Appetitive Pavlovian conditioning takes the past experience of rewards to form predictions and provide information about rewards. The bell in Pavlov's conditioning experiment has become a sausage predictor for the animal. By licking to the bell the dog demonstrates an expectation of the sausage that was evoked by the bell. Thus Pavlovian conditioning of the bell to the sausage has made the intrinsically neutral bell a reward predictor. The external stimulus or event (the bell) has become a predictor (of the sausage) and evokes an internal expectation (of the sausage) in the animal **(FIGURE 4)**. The same holds for any other reward predictor. The pineapple and the hamburger are predictors of the nutrients they contain.

Predictions tell us what is going to happen. This includes predictors of probabilistic rewards, even if the reward does not occur in every instance. Pavlov's bell predicts a sausage to the dog and at the same time induces salivation, licking, and approach. Thus Pavlovian conditioning confers two components, a predictive and an incentive property. The predictive component indicates what is going to happen now. The incentive property induces action, such as salivation, licking, and approach, which help to obtain and ingest the sausage. The two properties are separated in time in the classic and widely used delayed response tasks in which initial instructive stimuli have reward-predicting properties without eliciting a reward-directed action (but ocular saccades), whereas the final trigger or releasing stimulus induces the behavioral action and thus has both predictive and incentive properties. The predictive and incentive properties are separated spatially with different stimulus and goal (lever) positions and are dissociable in the behavior of particular rat strains. Specially bred sign-tracking rats approach the conditioned predictive stimulus, whereas goal trackers go directly to the reward upon stimulus appearance, indicating separation of predictive and incentive properties in goal trackers (163).

The predictive component can be further distinguished from an informational component. Once a stimulus has been Pavlovian conditioned, it confers information about the reward. The information does not necessarily predict explicitly what is actually going to happen every time, not even probabilistically. Through Pavlovian conditioning I have learned that a particular sign on a building indicates a pub because I have experienced the beer inside. The atmosphere and the beer represent a value to me that is assigned in a Pavlovian manner to the pub sign. I can pass the pub sign without entering the pub, however difficult that may be. Thus the sign is informational but does not truly predict a pint, only its potential, nor does it have the incentive properties at that moment to make me go in and get one. Then I may run into an unknown pub and experience a different beer, and I undergo another round of Pavlovian conditioning to the value of that particular pub sign. When I need to choose between different pubs, I use the information about their values rather than explicit predictions of getting a beer in every one of them. Thus Pavlovian conditioning sets up predictions that contain reward information. The predictions indicate that a reward is going to occur this time, whereas the reward informations do not necessarily result in reward every time.

The distinction between explicit prediction and information is important for theories of competitive decision-making. Value information about a reward is captured in the term of action value in machine learning, which indicates the value resulting from a particular action, without requiring to actually choosing it and obtaining that value (575). In analogy, object value indicates the value of a reward object irrespective of choosing and obtaining it. Individuals make decisions by comparing the values between the different actions or objects available and then selecting only the action or object with the highest value (see below **FIGURE 36*C***). Thus the decision is based on the information about

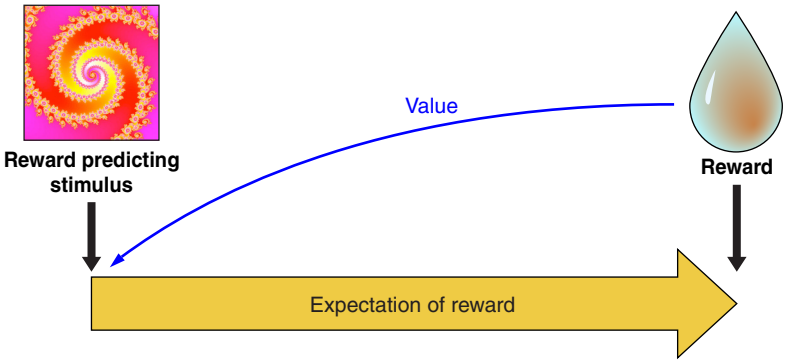


FIGURE 4. Pavlovian reward prediction. With conditioning, an arbitrary stimulus becomes a reward predictor and elicits an internal expectation of reward. Some of the behavioral reactions typical for reward occur also after the stimulus (Pavlovian stimulus substitution), in particular approach behavior, indicating that the stimulus has acquired reward value (blue arrow).

each reward value and not on the explicit prediction that every one of these rewards will be actually experienced, as only the chosen reward will occur. In this way, Pavlovian conditioning is a crucial building block for reward information in economic decisions. Separate from the decision mechanism, the acquisition and updating of action values and object values requires actual experience or predictions derived from models of the world, which is captured by model free and model-based reinforcement learning, respectively. Other ways to establish reward predictions and informations involve observational learning, instructions, and deliberate reflections that require more elaborate cognitive processes. All of these forms produce reward information that allows informed, competitive choices between rewards.

4. Pavlovian learning produces higher order rewards

Reward-predicting stimuli established through Pavlovian conditioning become higher order, conditioned rewards. By itself a red apple is an object without any intrinsic meaning. However, after having experienced its nutritious and pleasantly tasting contents, the apple with its shape and color has become a reinforcer in its own right. As a higher order, conditioned reward, the apple serves all the defining functions of rewards, namely, learning, approach behavior, and pleasure. The apple serves as a reinforcer for learning to find the vendor's market stand. When we see the apple, we approach it if we have enough appetite. We even approach the market stand after the apple has done its learning job. And seeing the delicious apple evokes a pleasant feeling. The apple serves these functions irrespective of explicitly predicting the imminent reception of its content or simply informing about its content without being selected. Thus Pavlovian conditioning labels arbitrary stimuli and events as higher order rewards that elicit all reward functions. All this happens without any physical change in the apple. The only change is in the eye of the beholder, which we infer from behavioral reaction.

The notion that Pavlovian conditioning confers higher order reward properties to arbitrary stimuli and events allows us to address a basic question. Where in the body is the original, unconditioned effect of rewarding substances located (517, 640)? The functions of alimentary rewards derive from their effects on homeostatic mechanisms involved in building and maintaining body structure and function. In the case of an apple, the effect might be an increase in blood sugar concentration. As this effect would be difficult to perceive, the attraction of an apple might derive from various stimuli conditioned to the sugar increase, such vision of the apple, taste, or other conditioned stimuli. Mice with knocked out sweet taste receptors can learn, approach, and choose sucrose (118), suggesting that the calories and the resulting blood sugar increase constitute the unconditioned reward effect instead of or in addition to taste. Besides being an unconditioned reward (by evoking pleasure or via innate mechanisms, see above), taste may also be a conditioned reward, similar to the sensory properties of other alimentary rewards, including temperature (a glass of unpleasant, luke warm water predicting reduced plasma osmolality) and viscosity (a boring chocolate drink predicting calories). These sensations guide ingestion that will ultimately lead to the primary reward effect. However, the conditioned properties are only rewarding as long as the original, unconditioned reward actually occurs. Diets lacking just a single essential amino acid lose their reward functions within a few hours or days and food aversion sets in (181, 239, 480). The essential amino acids are detected by chemosensitive neurons in olfactory cortex (314), suggesting a location for the original effects of amino acids with rewarding functions. Thus the better discernible higher order rewards facilitate the function of primary rewards that have much slower and poorly perceptible vegetative effects.

A similar argument can be made for rewards that do not address homeostasis but are based on the pleasure they evoke. The capacity of some unconditioned rewards to evoke pleasure and similar positive emotions is entirely determined by brain physiology. Pleasure as an unconditioned reward can serve to produce higher order, conditioned rewards that are also pleasurable. A good example are sexual stimuli, like body parts, that have unconditioned, innate reward functions and serve to make intrinsically neutral stimuli, like signs for particular stores or bars, predictors of sexual events and activity.

Taken together, the primary, homeostatic, or pleasurable reward functions are innate and determined by the physiology of the body and its brain that emerged from evolutionary selection. Individuals without such brains or with brains not sensing the necessary rewards have conceivably perished in evolution. The primary rewards come in various forms and "packages" depending on the environment in which individuals live. The same glucose molecule is packaged in sugar beets or bananas in different parts of the world. Conditioning through the actual experience within a given environment facilitates the detection of the primary rewards. Thus, in producing higher order rewards, Pavlovian learning allows individuals to acquire and optimally select homeostatic and pleasurable rewards whose primary actions may be distant and often difficult to detect.

5. Operant learning

Getting rewards for free sounds like paradise. But after Adam took Eve's apple, free rewards became more rare, and Pavlovian learning lost its exclusiveness. We often have to do something to get rewards. Learning starts with a reward that seems to come out of nowhere but actually came from an action, and we have to try and figure out what made it appear. The rat presses a lever by chance and receives a drop of sugar solution. The sugar is great, and it presses again, and again. It comes back for more. The frequency of its

behavior that results in the sugar increases. This is positive reinforcement, the more reward the animal gets the more it acts, the essence of Thorndike's Law of Effect (589). Crucial for operant learning is that the animal learns about getting the sugar drop only by pressing the lever. Without lever pressing it would not get any sugar, and the behavior would not get reinforced. If the sugar comes also without lever pressing, the sugar does not depend on lever pressing, and the rat would not learn to operate the lever (but it might learn in a Pavlovian manner that the experimental box predicts sugar drops).

Operant reinforcement means that we act more when we get more from the action. However, due to the somewhat voluntary nature of the action involved, the inverse works also. The more we act, the more we get. Our behavior determines how much we get. We get something for an action, which is the bonus or effort function of reward. We can achieve something by acting. We have control over what we get. Knowing this can motivate us. Operant learning allows behavior to become directed at a rewarding goal. We act to obtain a reward. Behavior becomes causal for obtaining reward. Thus operant learning represents a mechanism by which we act on the world to obtain more rewards. Coconuts drop only occasionally from the tree next to us. But we can shake the tree or sample other trees, not just because the coconut reinforced this behavior but also because we have the goal to get more coconuts. In eliciting goal-directed action, operant learning increases our possibilities to obtain the rewards we need and thus enhances our chance for survival and reproduction.

6. Value updating, goal-directed behavior, and habits

Value informations for choices need to be updated when reward conditions change. For example, food consumption increases the specific satiety for the consumed reward and thus decreases its subjective value while it is being consumed. In addition, the general satiety evolving in parallel lowers the reward value of all other foods. Although the values of all rewards ever encountered could be continuously updated, it would take less processing and be more efficient to update reward values only at the time when the rewards are actually encountered and contribute to choices. Reward values that are computed relative to other options should be updated when the value of any of the other option changes.

Operant learning directs our behavior towards known outcomes. In goal-directed behavior, the outcome is represented during the behavior leading to the outcome, and furthermore, the contingency of that outcome on the action is represented (133). In contrast, habits arise through repeated performance of instrumental actions in a stereotyped fashion. Habits are not learned with a new task from its outset but form gradually after an initial declarative, goal-directed learning phase (habit formation rather than habit learning). They are characterized by stimulus-response (S-R) performance that becomes "stamped-in" by reinforcement and leads to inflexible behavior. Despite their automaticity, habits extinguish with reinforcer omission, as do other forms of behavior. In permitting relatively automatic performance of routine tasks, habits free our attention, and thus our brains, for other tasks. They are efficient ways of dealing with repeated situations and thus are evolutionary beneficial.

Value updating differs between goal-directed behavior, Pavlovian predictions, and habits (133). Specific tests include devaluation of outcomes by satiation that reduces subjective reward value (133). In goal-directed behavior, such satiation reduces the operant response the next time the action is performed. Crucially, the effect occurs without pairing the action with the devalued outcome, which would be conventional extinction. The devaluation has affected the representation of the outcome that is being accessed during the action. In contrast, habits continue at unreduced magnitude until the devalued outcome is experienced after the action, at which point conventional extinction occurs. Thus, with habits, values are only retrieved from memory and updated at the time of behavior. Devaluation without action-outcome pairing is a sensitive test for distinguishing goal-directed behavior from habits and is increasingly used in neuroscience (133, 399, 646). Like goal-directed behavior, Pavlovian values can be sensitive to devaluation via representations and update similarly without repairing, convenient for rapid economic choices (however, Pavlovian conditioning is not goal "directed" as it does not require actions). Correspondingly, reward neurons in the ventral pallidum start responding to salt solutions when they become rewarding following salt depletion, even before actually experiencing the salt in the new depletion state (591). In conclusion, whether updating is immediate in goal-directed behavior or gradually with habit experience, values are only computed and updated at the time of behavior.

B. Neuronal Signals for Reward Learning

1. Basic conditions of learning: contiguity and contingency

In both Pavlovian and operant conditioning, an event (stimulus or action) is paired in time, and often in location, with a reinforcer to produce learning. In delay conditioning, the stimulus or action lasts until the reinforcer occurs. In trace conditioning, which is often less effective (47, 423), the stimulus or action terminates well before the reinforcer. Trace conditioning, but not delay conditioning, with aversive outcomes is disrupted by lesions of the hippocampus (47, 356, 553), although presenting the stimulus briefly again with the outcome restores trace conditioning (27). Thus temporal contiguity is important for conditioning.

What Pavlov did not know was that stimulus-reward pairing was not sufficient for conditioning. Rescorla showed in his "truly random" experiment that the key variable underlying conditioning is the difference in outcome between the presence and absence of a stimulus (or action) (459). Conditioning occurs only when reinforcement is dependent or "contingent" on the stimulus (or an action), and this condition applies also to rewards (126). When the same reward occurs also without the stimulus, the stimulus is still paired with reward but carries no specific reward information, and no conditioning occurs. Thus contingency refers to the intuitive notion that we learn only something that carries specific information or that decreases the uncertainty in the environment (reduction of informational entropy). Contingency is usually defined as difference in reinforcement between stimulus (or action) presence and background (stimulus absence) (**FIGURE 5*A***). A positive difference produces excitatory conditioning (positive learning, left of diagonal), whereas a negative difference produces inhibitory conditioning (negative learning, right of diagonal). Zero contingency produces no conditioning (diagonal line) (but likely perceptual learning). Backward conditioning (reward before stimulus) is ineffective or even produces inhibitory conditioning because the reward occurs without relation to the stimulus. Timing theories of conditioning define contingency as stimulus-to-intertrial ratio of reinforcement rates (175). More frequent reward during the stimulus compared with the no-stimulus intertrial interval makes the reward more contingent on the stimulus. When the same reward occurs later after stimulus onset (temporal discounting), stimulus-to-intertrial reward ratio decreases and thus lowers contingency, although other explanations are also valid (see below). Thus, after contiguity and stimulus-outcome pairing, contingency is the crucial process in conditioning. It determines reward predictions and informations, which are crucial for efficient decisions, and it produces higher order rewards, which drive most of our approach behavior.

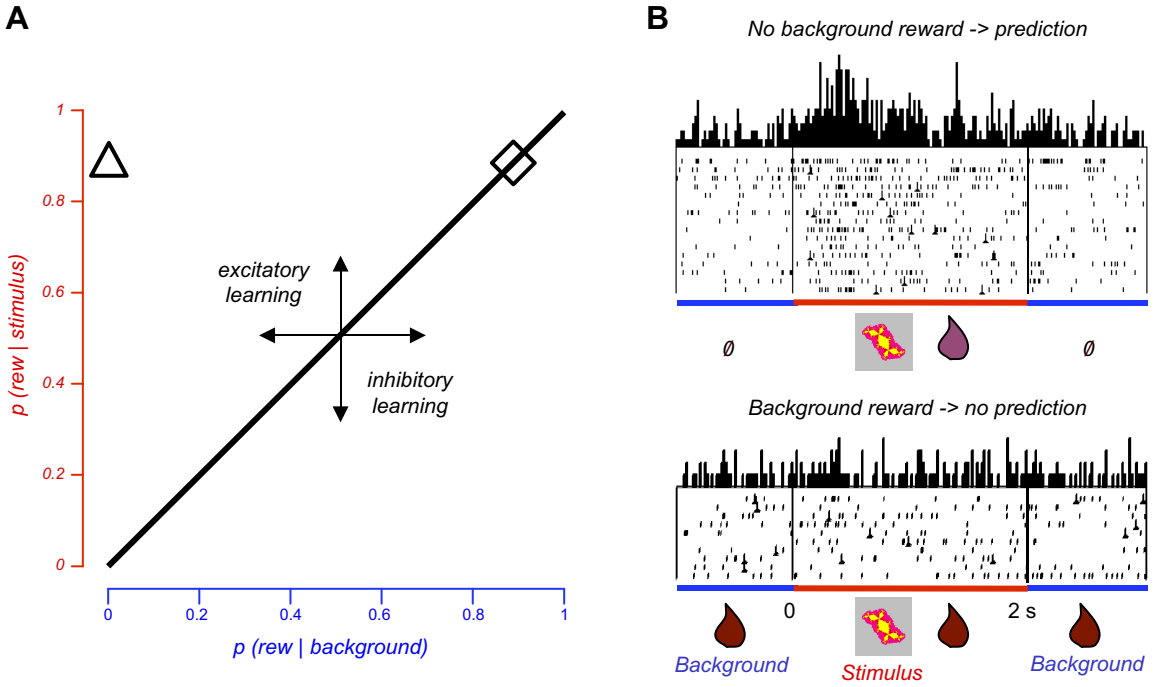


FIGURE 5. Reward contingency. *A:* role of contingency in learning. Contingency is shown as reward difference between stimulus presence and absence (background). Abscissa and ordinate indicate conditional reward probabilities. Higher reward probability in the presence of the stimulus compared with its absence (background) induces positive conditioning (positive contingency, triangle). No learning occurs with equal reward probabilities between stimulus and background (diagonal line, rhombus). Reward contingency applies to Pavlovian conditioning (shown here; reward contingent on stimulus) and operant conditioning (reward contingent on action). [Graph inspired by Dickinson (132).] *B:* contingency-dependent response in single monkey amygdala neuron. *Top*: neuronal response to conditioned stimulus (reward $P = 0.9$; red) set against low background reward probability ($P = 0.0$; blue) (triangle in *A*). *Bottom*: lack of response to same stimulus paired with same reward ($P = 0.9$) when background produces same reward probability ($P = 0.9$) (rhombus in *A*) which sets reward contingency to 0 and renders the stimulus uninformative. Thus the neuronal response to the reward-predicting stimulus depends entirely on the background reward and thus reflects reward contingency rather than stimulus-reward pairing. A similar drop in neuronal responses occurs with comparable variation in reward magnitude instead of probability (41). Perievent time histograms of neuronal impulses are shown above raster displays in which each dot denotes the time of a neuronal impulse relative to a reference event (stimulus onset, time = 0, vertical line at left; line to right indicates stimulus offset). [From Bermudez and Schultz (41).]

2. Neuronal contingency tests

Standard neuronal learning and reversal studies test contiguity by differentially pairing stimuli (or actions) with reward. Although they sometimes claim contingency testing, the proper assessment of contingency requires the distinction against contiguity of stimulus-reward pairing. As suggested by the "truly random" procedure (459), the distinction can be achieved by manipulating reward in the absence of the stimulus (or action).

A group of amygdala neurons respond to stimuli predicting more frequent or larger rewards compared with background without stimuli (**FIGURE 5*B***, *top*). These are typical responses to reward-predicting stimuli seen for 30 years in all reward structures. However, many of these neurons lose their response in a contingency test in which the same reward occurs also during the background without stimuli, even though stimulus, reward during stimulus and stimulus-reward pairing are unchanged (41) (**FIGURE 5*B***, *bottom*). Thus the responses do not just reflect stimulus-reward pairing but depend also on positive reward contingency, requiring more reward during stimulus presence than absence (however, amygdala neurons are insensitive to negative contingency). Although the stimulus is still paired with the reward, it loses its specific reward information and prediction; the amygdala responses reflect that lost prediction. Correspondingly, in an operant test, animals with amygdala lesions fail to take rewards without action into account and continue to respond after contingency degradation (399), thus showing an important role of amygdala in coding reward contingency.

Reward contingency affects also dopamine responses to conditioned stimuli. In an experiment designed primarily for blocking (see below **FIGURE 8*A***), a phasic environmental stimulus constitutes the background, and the crucial background test varies the reward with that environmental stimulus. Dopamine neurons are activated by a specific rewarded stimulus in the presence of an unrewarded environmental stimulus (positive contingency) but lose the activation in the presence of a rewarded environmental stimulus (zero contingency) (618). The stimulus has no specific information and prediction when the environmental stimulus already fully predicts the reward. Another experiment shows that these contingency-dependent dopamine responses may be important for learning. Using optogenetic, "rewarding" stimulation of dopamine neurons, behavioral responses (nose pokes) are only learned when the stimulation was contingent on the nose pokes, whereas contingency degradation by unpaired, pseudorandom stimulation during nose pokes and background induces extinction [truly random control with altered p(stimulation|nose poke)] (641).

Taken together, amygdala and dopamine neurons take the reward in the absence of the stimulus into account and thus are sensitive to contingency rather than simple stimulus-reward pairing. They fulfill the crucial requirement for coding effective reward information and prediction and thus may provide crucial inputs to neuronal decision mechanisms. Also, these neuronal signals demonstrate a physical basis for the theoretical concept of contingency and thus strengthen the plausibility of the most basic assumption of animal learning theory.

3. Prediction error learning theory

Contingency, in addition to stimulus-reward pairing (contiguity), is necessary for rewards to induce conditioning. Reward is contingent on a stimulus only when it differs between stimulus presence and absence. When reward is identical, including its time of occurrence, between stimulus presence and absence, it is already predicted by the objects that exist during stimulus presence and absence and therefore not surprising. Hence, the stimulus will not be learned.

The formal treatment of surprise in conditioning employs the concept of prediction error. A reward prediction error PE is the difference between received reward λ and reward prediction V in trial t

$$\mathrm{PE}(t) = \lambda(t) - \mathrm{V}(t) \qquad (1)$$

This formal definition of "error" extends beyond the colloquial meaning of inaccurate behavior. Prediction errors occur whenever *Equation 1* applies, irrespective of whether they simply occur during behavioral learning or are actually being used for learning.

Animal learning theory aims to explain the role of contingency in conditioning by formalizing how prediction errors update reward predictions from previously experienced rewards (460). The new prediction V for trial $t+1$ derives from the current prediction V(t) and the prediction error PE(t) weighted by learning rate α

$$\mathrm{V}(t+1) = \mathrm{V}(t) + \alpha * \mathrm{PE}(t) \qquad (2)$$

Note that V captures the sum of predictions if several stimuli are present. According to this account, conditioning constitutes error driven learning **(FIGURE 6*A*)**. When the outcome differs from the prediction, a prediction error proportional to that difference is generated and added to the current prediction after weighting by learning rate α. When α is set to 1.0, the new prediction is updated in one trial. However, learning rates are usually <1.0, which leads to the typical asymptotic learning curves during which prediction errors decline gradually **(FIGURE 6*B*)**.

Rewards that are better than predicted generate positive prediction errors and lead to response acquisition, whereas worse rewards generate negative prediction errors and lead to extinction of behavior. When the prediction error becomes zero, no further learning occurs and the prediction

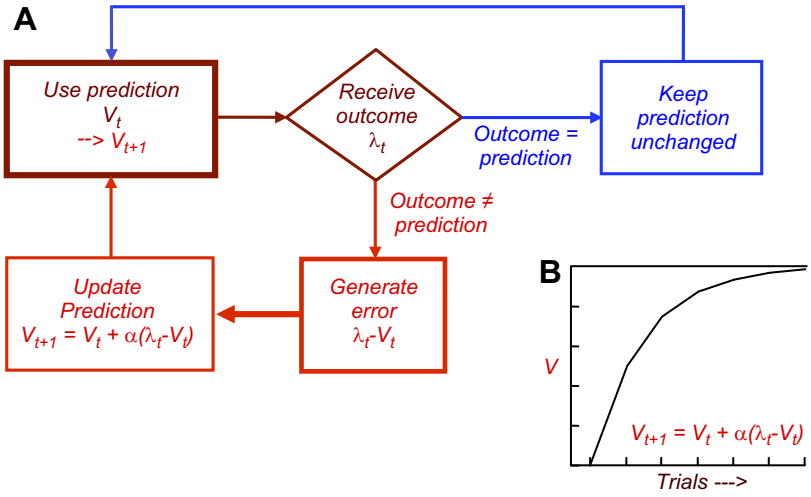


FIGURE 6. Learning with prediction errors. *A:* feedback circuit diagram for prediction updating by error. An error is generated when outcome (reward, punisher) differs from its prediction. In Pavlovian conditioning, a prediction error after an outcome change leads to prediction updating which leads to a behavioral change. In operant conditioning, a prediction error after an outcome change leads to a behavioral change which leads to prediction updating. In contrast, no prediction error is generated when the outcome matches the prediction, and behavior remains unchanged. V is reward prediction, λ is reward, α is learning rate, and *t* is trial. *B:* typical learning curve, generated by gradually declining prediction errors [λ(*t*) − V(*t*)].

remains stable. Thus learning advances only to the extent to which a reward is unpredicted and slows progressively with increasingly accurate predictions. This formalism views learning intuitively as a change in behavior that occurs when encountering something new or different than predicted, whereas behavior stays the same when everything occurs according to "plan."

Positive prediction errors direct organisms to maximal rewards, whereas negative prediction errors direct them away from poor outcomes. Obviously, individuals appreciate positive prediction errors and hate negative ones. Optimal behavior for maximizing rewards is characterized by many positive prediction errors and few negative ones.

Contingency is based on prediction errors. Contingency is non-zero, and prediction errors are elicited, when reward differs between stimulus presence and absence. This is the crucial requirement for learning. When reward is identical during stimulus presence and absence, there is no contingency, no prediction error, and no learning.

Specific learning situations benefit from extensions of the basic Rescorla-Wagner rule. Attentional learning rules relate the capacity to learn (associability) to the degree of attention evoked by the stimulus, which depends on degree of surprise (prediction error) experienced in the recent past (prediction error in the current trial cannot contribute to attention to the stimulus, as the error occurs only after the stimulus). Learning rate α is modified according to the weighted unsigned (absolute) prediction error in the recent past (425)

$$\alpha = \kappa \left| \lambda(t) - V(t) \right| \qquad (3)$$

with κ as weighting parameter. Thus learning rate increases with large prediction errors, as in the beginning of learning or with large step changes, resulting in strong influences of prediction error and thus fast learning according to *Equation 2*. In contrast, learning slows with small prediction errors from small changes.

When rewards are drawn from different probability distributions, comparable learning can be achieved by scaling the update component α*PE to the standard deviation σ of the respective reward distribution by either (438)

$$\alpha^{\circ} = \alpha / \sigma \qquad (3A)$$

or

$$PE(t)^{\circ} = PE(t) / \sigma \qquad (3B)$$

Behavioral tests with measurable prediction errors demonstrate that scaling learning rate α (*Equation 3A*) by standard deviation may produce constant learning rates with probability distributions differing only in standard deviation (384), which can be formulated as

$$\alpha = \kappa / \sigma \left| \lambda(t) - V(t) \right| \qquad (3C)$$

Thus risk may affect learning in two separate ways, by increasing learning rate α via the attentional effect of the prediction error (*Equation 3*), and by decreasing learning rate or prediction error via division by standard deviation σ (*Equations 3A* and *3B*).

The Rescorla-Wagner learning rule models Pavlovian conditioning: changes in outcome lead to prediction errors that lead to changes in prediction and consequently changes in behavior. The learning rule extends to operant conditioning: changes in outcome produce prediction errors that lead to behavioral changes which then result in changes in prediction.

In realistic situations, rewards may be unstable, which results in frequent prediction errors, continuous updating of predictions, and correspondingly varying behavior. Thus learning of a single, stable reward is a special case of, and

formally equivalent to, the general process of updating of reward predictions.

Although the Rescorla-Wagner rule models the acquisition of predictions (V in *Equation 2*), it does not conceptualize predictions as conditioned, higher order reinforcers. However, higher order reward properties are responsible for most reward functions in learning, approach behavior, and emotional reactions, as outlined above (640). The temporal difference (TD) reinforcement model in machine learning captures the learning functions of higher order rewards (574). Its prediction error includes both unconditioned and higher order reinforcers, without distinguishing between them, as

$$\mathrm{TDPE}(t) = [\lambda(t) + \gamma\sum \mathrm{V}(t)] - V(t-1) \quad (4)$$

The t stands now for time step within a given trial, γ is a temporal decay (discounting) factor. A temporal, rather than trial by trial, prediction error is adequate when including higher order reinforcers that derive from reward predictors and necessarily occur at earlier times than unconditioned reinforcers. The term $[\lambda(t) + \gamma\sum \mathrm{V}(t)]$ replaces the $\lambda(t)$ term in *Equation 1* and incorporates at every time step t the unconditioned reward $\lambda(t)$ and the discounted (γ) sum ($\sum$) of future rewards predicted by $\mathrm{V}(t)$ serving as higher order reward. The TDPE includes a higher order prediction error that indicates how much the reward predicted by a conditioned stimulus varies relative to an earlier reward prediction. Thus TDPEs are generated by successive conditioned stimuli to the extent that these stimuli carry different reward predictions than the preceding stimulus.

Substituting PE of *Equation 2* by TDPE of *Equation 4* leads to the TD learning model

$$\mathrm{V}(t+1) = \mathrm{V}(t) + \alpha * \mathrm{TDPE}(t) \quad (5)$$

Thus TD learning advances backwards from an unconditioned reward to the earliest reward predictor. Updating of the earliest reward prediction employs the prediction value of the subsequent stimulus and does not need to wait for the final, unconditioned reward. Thus each of the conditioned, reward-predicting stimuli serves as a higher order reward, which is biologically plausible by reflecting the frequent occurrence of conditioned rewards and addressing the often difficult distinction between the primary or conditioned nature of natural rewards (640). The use of truly unconditioned rewards in TD learning is straightforward (573). TD reinforcement models are efficient and learn complex tasks involving sequential states, like balancing a pole on a moving platform (32) or playing backgammon (586).

4. Dopamine reward prediction error signal

Most midbrain dopamine neurons show rather stereotyped, phasic activations with latencies of <100 ms and durations of <200 ms following unpredicted food or liquid rewards **(FIGURE 7, *A* AND *B*)**. The response codes the prediction error, namely, the quantitative difference between received and predicted reward value in each trial t, which can be generically be expressed as

$$\mathrm{DaResp}(t) \sim \lambda(t) - \mathrm{V}(t) \quad (6)$$

The neuronal response is bidirectional, as a reward that is better than predicted elicits an activation (positive prediction error response) and a reward that is worse than predicted induces a depression (negative error response) in monkeys, rats, and mice **(FIGURE 7*C*)**. A fully predicted reward draws no response. These responses occur with all-or-none reward prediction errors (321, 365, 410, 503, 517, 521, 524), and with graded, quantitative prediction errors (more or less reward than predicted) (33, 102, 144, 161, 376, 598). Thus the dopamine response codes a reward prediction error (*Equation 1*) that may constitute a biological instantiation of the reinforcement term in Rescorla-Wagner learning (365), TD learning (374), and spectral timing (71).

The negative dopamine reward prediction error response consists of a depression of activity against low background activity. Adequate assessment requires measuring impulse rate over the maximal length of the depression (34), thus avoiding an almost ungraded negative response (33). Even with a graded negative response (161), the negative error response has less dynamic range than the positive error response. However, the complete cessation of dopamine activity for varying durations may have stronger impact on postsynaptic neurons than graded increases and thus may compensate for the limited dynamic range. Even before it comes to synaptic transmission, voltammetrically measured dopamine concentration changes resulting from negative reward prediction errors are symmetric to the changes resulting from positive errors (204). Thus the dopamine reward prediction error signal seems to be bidirectional.

Dopamine responses show features beyond the Rescorla-Wagner learning rule. About one-third of dopamine neurons show also slower activations preceding reward that vary with the standard deviation (risk) of reward probability distributions (161). The risk response reflects the unsigned (absolute) prediction error and, in keeping with the attentional learning rules (425), may affect learning rate α (*Equation 3*) and thus serve to adjust learning speed. Furthermore, the prediction error response itself scales with the standard deviation of reward probability distributions (598) according to *Equation 3B*, which may be driven by the slower activation coding standard deviation and serve to achieve comparable learning with different standard deviations. Third, the prediction error response is sensitive to predictions in model-based learning (see below).

The dopamine response at the time of reward fulfils stringent tests for prediction errors conceptualized by animal

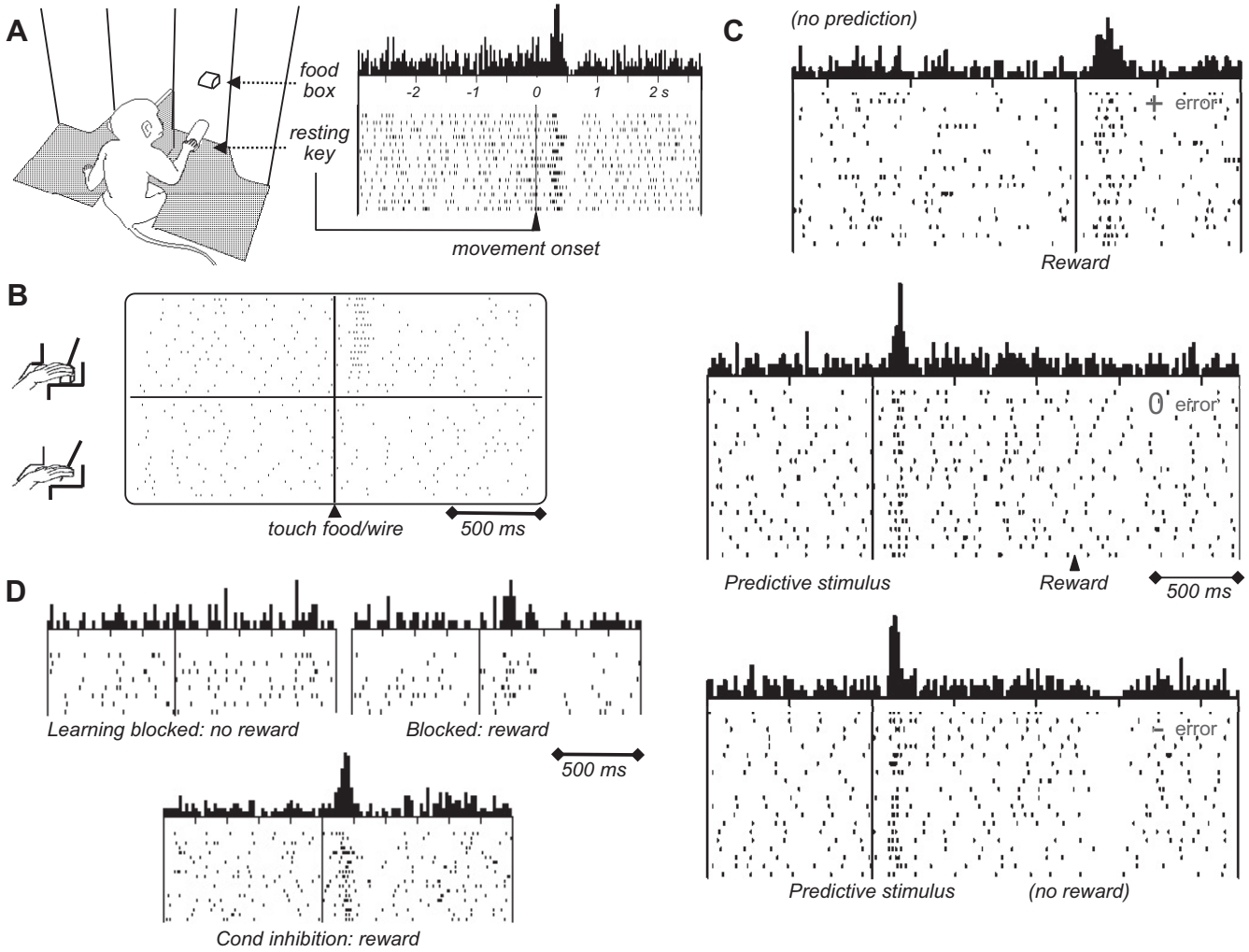


FIGURE 7. Dopamine prediction error responses at the time of reward in monkeys. *A:* dopamine responses to touch of food without any phasic stimuli predicting the reward. The food inside the box is invisible but is touched by the hand underneath the cover. Movement onset is defined as release of resting key. *B:* differential response to touch of a piece of apple (*top*) but not to touch of a bare wire (*bottom*) or known inedible objects. Left graphs show the animal's hand entering the covered food box. Inside the box, touch of a bare wire or a wire holding the food elicits an electric signal for temporal reference (vertical line at right). [*A* and *B* from Romo and Schultz (491), with kind permission from Springer Science and Business Media.] *C:* reward prediction error responses at time of reward (*right*) and reward-predicting visual stimuli (left in 2 bottom graphs). The dopamine neuron is activated by the unpredicted reward eliciting a positive reward prediction error (blue + error, *top*), shows no response to the fully predicted reward eliciting no prediction error (0 error, *middle*), and is depressed by the omission of predicted reward eliciting a negative prediction error (- errror, *bottom*). [From Schultz et al. (524).] *D:* reward prediction error responses at time of reward satisfy stringent prediction error tests. *Top*: blocking test: lack of response to reward absence following the stimulus that was blocked from learning (*left*), but activation by surprising reward after blocked stimulus (*right*). [From Waelti et al. (618).] *Bottom*: conditioned inhibition test. Supranormal activation to reward following an inhibitory stimulus explicitly predicting no reward. [From Tobler et al. (597).]

learning theory **(FIGURE 7*D*)**. In the blocking test, a test stimulus is blocked from acquiring reward prediction when the reward is fully predicted by another stimulus (zero contingency). If reward does occur after the nonpredictive test stimulus, it will produce a positive prediction error, and dopamine neurons are activated by that reward (618). In the conditioned inhibition test, a reward occurring after a stimulus explicitly predicting reward absence elicits a super strong positive reward prediction error (because of a negative prediction being subtracted from the reward, *Equation 1*), and dopamine neurons accordingly show supranormal activation by the surprising reward (*Equation 6*) (598). Thus the phasic dopamine responses follow the formal theoretical requirements for prediction error coding.

Reward-predicting stimuli induce phasic activations in most dopamine neurons (60-75%) **(FIGURE 7*C*)** (322, 503) and increase correlations between dopamine impulses (256). These responses are gradually acquired through

learning and thus code the prediction V of Rescorla-Wagner and TD models. Correspondingly, voltammetrically measured striatal dopamine release shifts from reward to reward-predicting stimuli during conditioning (117). Acquisition of dopamine responses to conditioned stimuli requires prediction errors, as the blocking test shows. Compounding a novel stimulus with an established reward predictor without changing the reward fails to elicit a prediction error and prevents the stimulus from being learned **(FIGURE 8*A*)**. A conditioned inhibitor elicits a depressant instead of an activating response **(FIGURE 8*B*)**.

Results from tests with sequential reward-predicting stimuli suggest that dopamine responses signal higher order reward prediction errors compatible with TD learning (374). Thus the dopamine responses are well described by extending *Equation 6* and using *t* as time steps, which can be generically be expressed as

$$\mathrm{DaResp}(t) \sim [\lambda(t) + \gamma\sum \mathrm{V}(t)] - \mathrm{V}(t-1) \quad (7)$$

The dopamine response jumps sequentially backwards from the "unconditioned" liquid reward via stimuli predicting this reward to the earliest reward-predicting stimulus, while losing the response to the predicted intermediate stimuli and the predicted reward **(FIGURE 8*C*)** (453, 521). Similar to TD learning, the dopamine responses do not distinguish between primary and higher order reward prediction errors. Only their magnitudes, and fractions of responding neurons, decrease due to temporal discounting, which in TD learning is incorporated as γ. With several consecutive conditioned stimuli, dopamine neurons code the TD prediction error exactly as the predicted value changes between the stimuli (398). In such tasks, dopamine responses match closely the temporal profiles of TD errors that increase and decline with sequential variations in the discounted sum of future rewards **(FIGURE 8*D*)** (144). Thus dopamine neurons code reward value at the time of conditioned and unconditioned stimuli relative to the reward prediction at that moment.

The reward-predicting instruction stimuli in delay tasks have little incentive properties and thus provide a distinction between predictive and incentive stimulus properties (see above). These instructions contain information and elicit saccadic eye movements for acquiring the information contained in the stimulus, but the animal is not allowed to react immediately with a movement towards a rewarded target; it must wait for a later, incentive, movement-triggering stimulus. The consistent dopamine responses to these predictive instructions **(FIGURE 8*C*)** (301, 349, 453, 521, 527, 579), and their lack of ocular relationships (527), suggest predominant coding of predictive stimulus properties, without requiring incentive properties. In contrast, in specific sign- and goal-tracking rat strains, striatal voltammetric dopamine responses are stronger to stimuli with incentive compared with reward-predictive properties (163). These differences may be due to different methods (electrophysiological impulses vs. voltammetric dopamine concentrations) and genetic differences in the specifically bred sign versus goal trackers differentially affecting the dopamine response transfer from reward to predictive stimulus.

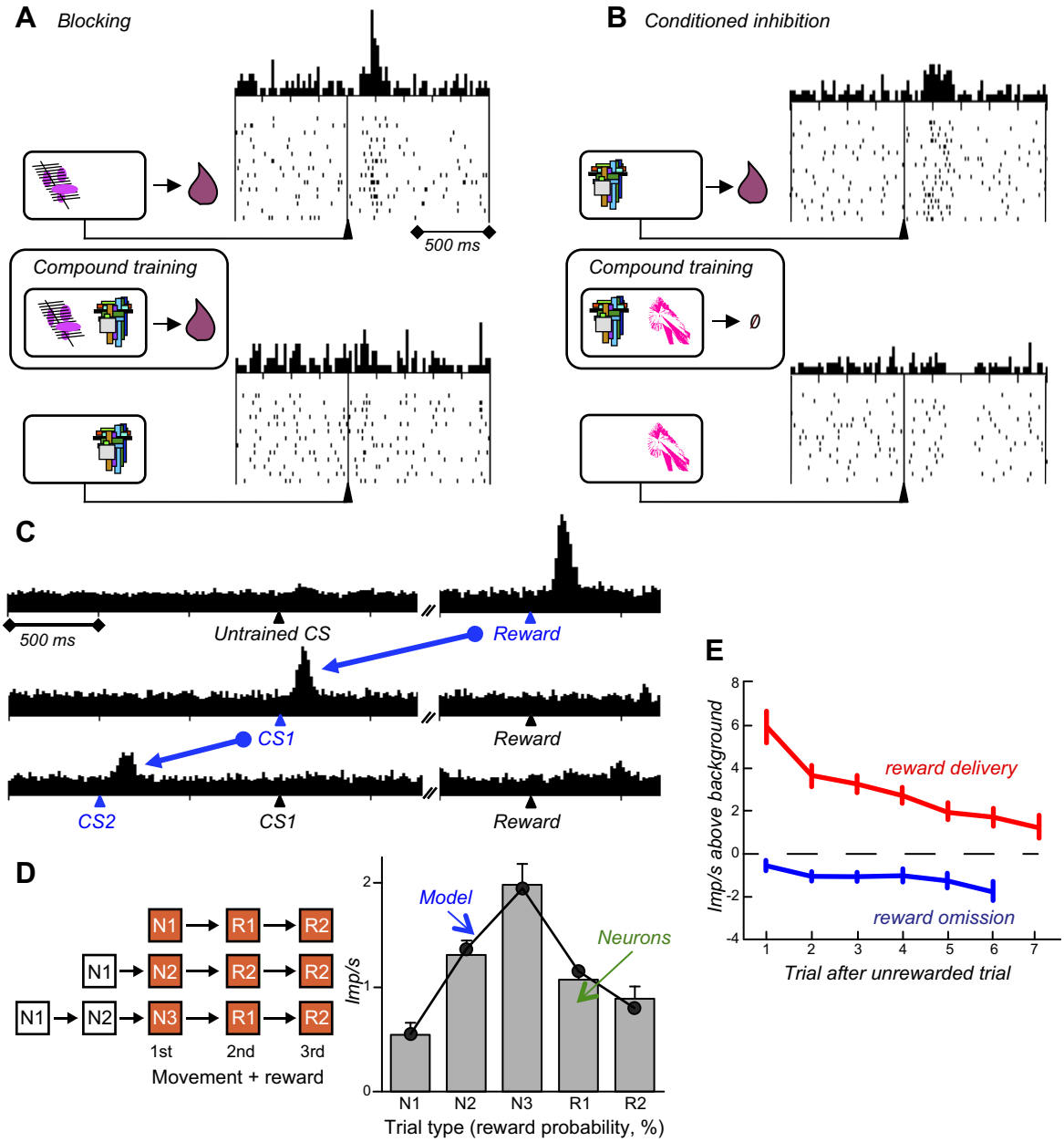
Hundreds of human neuroimaging studies demonstrate activations induced by reward related stimuli and reward-related actions in the main reward structures (280, 590). A BOLD neuromagnetic (fMRI) signal reflecting reward prediction error is found in the ventral striatum (351, 390) and constitutes probably the most solid reward response in the brain. It likely reflects the dopamine prediction error signal, as it increases with dopamine agonists and decreases with dopamine antagonists (431). It is detected more easily in striatal and frontal dopamine terminal areas than in midbrain cell body regions (111, 431, 615), because it presumably reflects summed synaptic potentials (328), which may be stronger after the massive axon collaterilization in the striatum. The human reward signal allows also to assess neuronal correlates of positive emotional reward functions, which is intrinsically difficult in animals despite recent efforts (44). With the use of raclopride binding with positron emission tomography (PET), amphetamine-induced dopamine concentrations correlate highly with subjective euphoria ratings in ventral (but not dorsal) striatum (140). BOLD responses in ventral striatum correlate with pleasantness ratings for thermal stimuli (along with orbitofrontal and anterior cingulate activations) (486) and with momentary happiness ratings derived from a combination of certain rewards, gambles, and reward prediction errors (along with anterior cingulate and anterior insula activations) (498). These responses in the dopamine receiving striatum suggest a role of dopamine in signaling pleasure and happiness.

5. Dopamine error signal incorporates model-based predictions

Standard conditioning occurs through prediction errors derived from actually experienced outcomes. However, individuals also learn about contexts, situations, rules, sequences, and conditional probabilities, which may be summarily called models of the world. Knowledge derived from such models can affect the prediction in the neuronal prediction error computation, result in a more appropriate teaching signal, and thus tremendously improve learning.

Model-based learning involves two separate processes, the acquisition and updating of the model, and the influence of the model on reinforcement learning, which likely occurs by influencing predictions of the outcome (139, 575). The acquisition and updating of some of the models likely involves cortical rather than dopamine signals. In contrast, dopamine responses seem to incorporate the predictive information from models once they are established.

In one such experiment, sequences of unrewarded trials lead to higher conditional reward probability with increasing trials (increasing hazard function). The underlying temporal structure constitutes a model in which reward prediction increases progressively. Due to increasing reward prediction, later reward delivery induces increasingly weaker positive prediction errors, and reward omission elicits stronger negative prediction errors. In contrast, model-free reinforcement learning would only consider the past unrewarded trials and hence generate progressively decreasing reward prediction and an opposite pattern of errors. In this experiment, dopamine prediction error responses decrease with increasing numbers of trials since the last reward and thus reflect the increasing reward prediction **(FIGURE 8*E*)** (382). In another experiment, dopamine reward prediction error responses adapt to previously learned reward probability distributions (598). The responses scale to expected value and standard deviation and show similar magnitudes with 10-fold differences in reward magnitude. In another example of model-based learning, acquisition of a reversal set allows animals to infer reward value of one stimulus after a single reversed trial with the other stimulus. Accordingly, dopamine prediction error responses reflect the inferred reversed reward prediction (68). In each of these examples, the observed dopamine prediction error responses are not fully explained by the immediate experience with reward but incorporate predictions from the task structure (model of the world). Thus dopamine neurons process prediction errors with both model free and model-based reinforcement learning.



6. Dopamine response in cognitive tasks

Bidirectional dopamine prediction error responses occur also in tasks with elaborate cognitive components. Dopamine neurons, and their potential input neurons in lateral habenula, show bidirectional prediction error responses that signal values derived from differential discounting with conditioned reinforcers (66, 67), although concurrent risk influences on subjective value need to be ruled out. More complex tasks require distinction of many more possible confounds. Early studies used spatial delayed, spatial alternation, and delayed matching-to-sample tasks typical for investigating neurons in prefrontal cortex and striatum and report positive and negative prediction error responses to rewards and reward-predicting stimuli compatible with TD learning models (321, 521, 579). Similar dopamine error responses occur in tasks employing sequential movements (503), random dot motion (389), and somatosensory detection (122). Advanced data analyses from these complex tasks reveal inclusion of model-based temporal predictions into the dopamine error response (382) and compatibility with TD models (144). Variations in stimulus coherence (389) and visual search performance (349) result in graded reward probabilities that induce reward prediction errors that are coded by dopamine neurons, although the error coding may not be easily apparent (426). Thus, in all these high-order cognitive tasks, the phasic dopamine response tracks only, and reliably, the reward prediction error.

7. Dopamine event detection response

The dopamine reward prediction error response is preceded by a brief activation that detects any sufficiently strong event before having identified its reward value (**FIGURE 9, *A–C***, blue areas). The early response component occurs with latencies of <100 ms with all sensory modalities, including loud noises, light flashes, and rapidly moving visual stimuli (229, 322, 527, 563) and evokes dopamine release (117). The activation reflects the sensory impact and physical salience of rewards, punishers, and their predictive stimuli (157, 159, 160). It does not fully reflect motivational salience, as it is absent with negative motivating events of low sensory impact (160, 366), negative reward prediction errors (321, 517, 521, 524, 597, 618), and conditioned inhibitors (597). The detection response covaries inversely with the hazard rate (of temporal event occurrence) and thus codes surprise salience (286) or a temporal prediction error (389) as conceptualized in TD models (574). Thus the early dopamine response component reflects the detection of an event before having identified its value.

This initial response component develops into the graded reward prediction error response, which ranges from depression with strong negative prediction errors to strong activation with positive prediction errors (158, 285, 301, 376, 597). The transition from the first to the second response component depends on the time necessary for assessing the value of the stimulus and thus may vary considerably between behavioral tasks. With easy identification, the transfer takes ~50 ms and fuses the two components, for example, when fixed stimulus positions allow rapid stimulus identification without eye saccades (**FIGURE 9, *A* AND *B***). Then the two components are fused and may appear as a single response whose identity has long been unclear and may occasionally be labeled as attentional (355, 356). A possible intermediate sensory identification step is either not discernible or is not engaging dopamine neurons. In contrast, the components separate into two separate responses with slower stimulus identification, as seen with random dot motion stimuli requiring >200 ms behavioral discrimination time (**FIGURE 9*C***) (389). Thus the early dopamine activation detects events rapidly before identifying them properly, whereas the subsequent bidirectional response component values the event and codes a proper reward prediction error.

The accurate value coding by the second response component is reflected in the error response at the time of the reward (398, 618). Despite the initial dopamine activation by an unrewarded stimulus, a subsequent reward is registered as a surprise eliciting a positive prediction error re-

FIGURE 8. Dopamine responses to conditioned stimuli in monkeys. *A*: stimulus responses of a single dopamine neuron during a blocking test. A pretrained stimulus predicts liquid reward and induces a standard dopamine response (*top*). During compound training, a test stimulus is shown together with the pretrained stimulus while keeping the reward unchanged (*middle left*). Thus the reward is fully predicted by the pretrained stimulus, no prediction error occurs, and the test stimulus is blocked from learning a reward prediction. Correspondingly, the test stimulus alone fails to induce a dopamine response (*bottom*). [From Waelti et al. (618).] *B:* stimulus responses of a single dopamine neuron during a conditioned inhibition test. The reward normally occurring with the pretrained stimulus (*top*) fails to occur during compound training with a test stimulus (*middle left*). This procedure makes the test stimulus a predictor of no reward which correspondingly induces a dopamine depression (*bottom*). [From Tobler et al. (597).] *C:* stepwise transfer of dopamine response from reward to first reward-predicting stimulus, corresponding to higher order conditioning conceptualized by temporal difference (TD) model. CS2: instruction, CS1: movement releasing (trigger) stimulus in delayed response task. [From Schultz et al. (521).] *D:* reward prediction error responses closely parallel prediction errors of formal TD model. *Left*: sequential movement-reward task. One or two stimuli (white N1, N2) precede the movements leading to reward (orange, 1st, 2nd, 3rd). *Right*: averaged population responses of 26 dopamine neurons at each sequence step (gray bars, numbers indicate reward probabilities in %) and time course of modeled prediction errors $\{[\lambda(t) + \gamma\sum V(t)] - V(t-1)\}$ (black line). [From Enomoto et al. (144).] *E:* dopamine prediction error responses reflect model-based reward prediction derived from temporal task structure. The model specifies an increase in conditional reward probability P(reward | no reward yet) from initial $P = 0.0625$ to $P = 1.0$ after six unrewarded trials. Correspondingly, positive prediction errors with reward occurrence decrease across successive trials, and negative errors with reward omission increase. Averaged population responses of 32 dopamine neurons show similar temporal profiles (blue and red, as impulses/s above neuronal background activity). [From Nakahara et al. (382), with permission from Elsevier.]

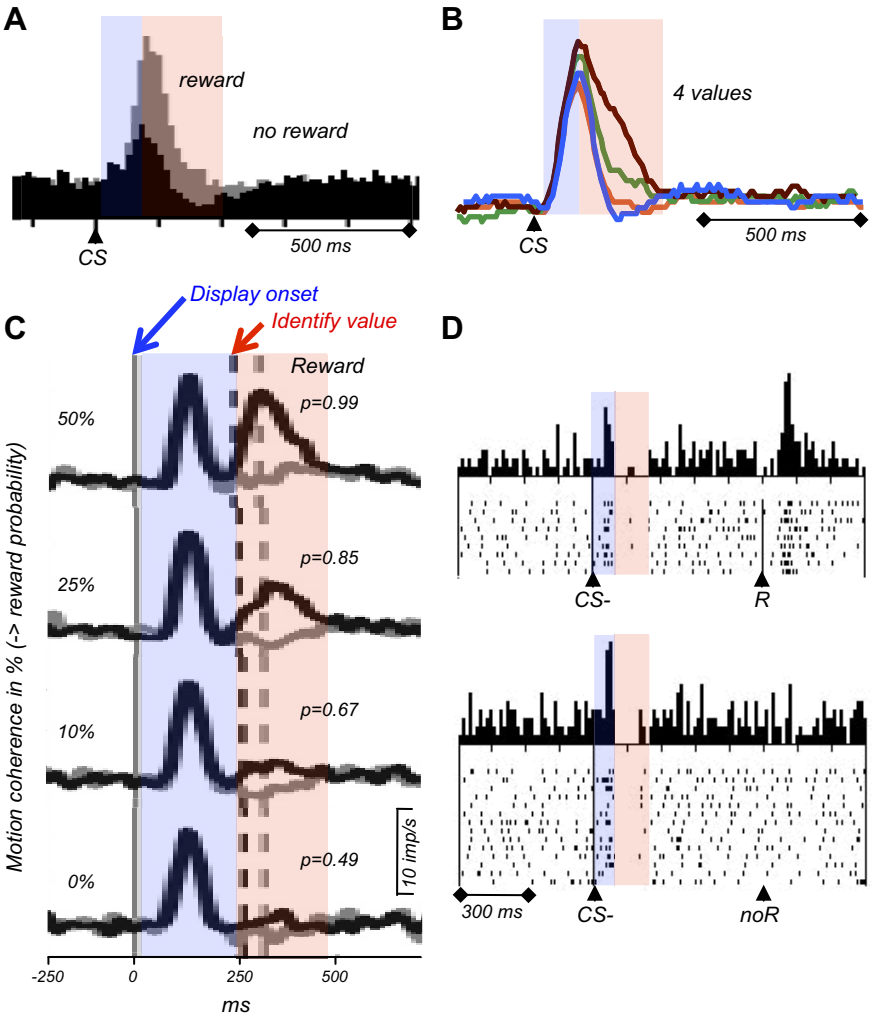


FIGURE 9. Two components of phasic dopamine responses. *A:* averaged population responses of 69 monkey dopamine neurons to conditioned stimuli (CS) predicting reward (gray) and no reward (black). Note the initial indiscriminate detection response component (blue) and the subsequent reward response component distinguishing between reward and no reward prediction (red). [From Tobler et al. (597).] *B:* averaged population responses of 54 monkey dopamine neurons to conditioned stimuli (CS) predicting rewards at different delays (2, 4, 8, and 16 s; brown, green, orange, and blue, respectively). The value reduction due to temporal discounting affects only the second, reward prediction error component (red). [From Kobayashi and Schultz (285).] *C:* differentiation of dopamine response into initial detection response and subsequent prediction error response. Increasing motion coherence (from 0 to 50%) improves binary dot motion discrimination and translates into increasing reward probability (from $P = 0.49$ to $P = 0.99$). The first response component is nondifferentially constant (blue), whereas the second component grows with increasing reward value (derived from probability, *bottom* to *top*, red). [From Nomoto et al. (389).] *D:* accurate value coding at time of reward despite initial indiscriminate stimulus detection response. After the unrewarded conditioned stimulus (CS-), surprising reward (R) elicits a positive prediction error response (*top*), whereas predicted reward absence (noR) fails to elicit a negative error response (*bottom*). [From Waelti et al. (618).]

sponse (**FIGURE 9*D***, *top*); accordingly, reward absence fails to induce a negative prediction error response (**FIGURE 9*D***, *bottom*). Thus, from the second response component on, and before a behavioral reaction can occur, dopamine responses accurately reflect the unrewarded stimulus nature. This constitutes the information postsynaptic neurons are likely to receive, which may explain why animals do not show generalized behavioral responses despite initial neuronal response generalization.

The initial dopamine detection response component varies with four factors. First, it increases with the sensory impact of any physical event, which confers physical salience (**FIGURE 10*A***) (160). This is most likely what the early studies on dopamine salience responses observed (229, 322, 563). Second, also involving memory processing stages, the dopamine detection component is enhanced by stimulus novelty (conferring novelty salience) (322) or surprise (conferring surprise salience) (255, 286, 345). Although novelty may constitute outright reward (458), dopamine neurons are not activated by novelty per se but require stimuli with sufficient sensory impact. The neurons do not respond to small novel stimuli, even though they show prediction error responses to the same small stimuli after learning (597, 618). The total dopamine response decreases with stimulus repetition, together with the animal's ocular orienting responses, and increases again with reward prediction learning (**FIGURE 10*B***) (517). Third, the dopamine detection component increases with generalization to rewarded stimuli, even when animals discriminate behaviorally well between them. The neuronal generalization is analogous to generalization in behavioral conditioning when an unconditioned stimulus resembles closely a conditioned stimulus and the "associative strength" spills over (335). The incidence of activations to unrewarded, including aversive, stimuli increases with closer similarity and same sensory modality as rewarded stimuli (527, 597, 618). A change from auditory to visual modality of the rewarded stimulus, while leaving the visual aversive stimulus unchanged, increases the initial activations from <15 to 65% of dopamine neurons (**FIGURE 10*C***) (366). Generalization likely also induces motivational salience via similar positively valued stimuli. Generalization may explain the more frequent dopamine activations to conditioned compared with primary aversive stimuli (345), which defies basic learning concepts. The initial response component evoked by reward

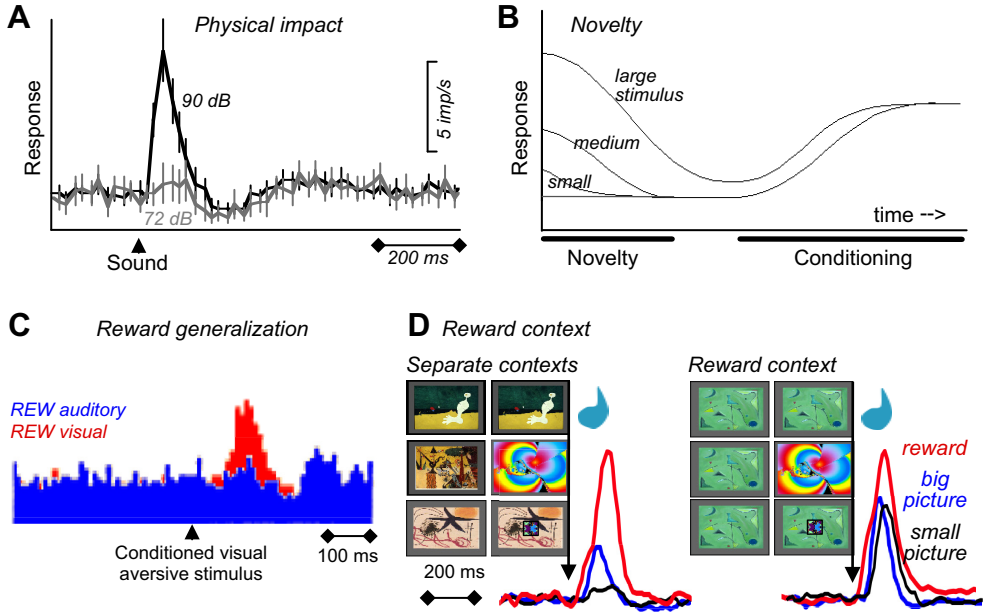


FIGURE 10. Four factors influencing the detection component of phasic dopamine responses. *A:* detection response generated by sensory impact, conferring physical salience. Louder, nonaversive sounds with higher physical salience generate stronger activations (72 and 90 dB, respectively; behavioral choice preferences demonstrated their nonaversive nature). Averaged population responses measured as impulses/s (imp/s) of 14 and 31 monkey dopamine neurons, respectively. [From Fiorillo et al. (160).] *B:* detection response, and possibly also second response component, enhanced by stimulus novelty, conferring novelty or surprise salience. Stimulus novelty itself is not sufficient to induce dopamine activations, as shown by response absence with small stimuli (horizontal line), but enhances detection response when stimuli are physically larger and more salient (vertical axis). Neuronal responses wane with stimulus repetition due to loss of novelty and increase again with conditioning to reward (from *left* to *right*). [Composite scheme from Schultz (517), derived from original data (221, 597, 618).] *C:* detection response enhanced by generalization to rewarded stimuli. Blue: minor population response to conditioned visual aversive stimulus alternating with auditory reward-predicting stimulus (REW auditory) (active avoidance task). Red: substantial activation to identical visual aversive stimulus when the alternate reward-predicting stimulus is also visual (REW visual), a situation more prone to stimulus generalization. As control, both auditory and visual reward-predicting stimuli induce typical dopamine activations (not shown). [From Mirenowicz and Schultz (366).] *D:* detection response enhanced by reward context. *Left* (separate contexts): minor dopamine population activations induced by unrewarded big and small pictures when non-reward context is well separated from reward context by testing in separate trial blocks, using distinct background pictures and removing liquid spout in picture trials. *Right* (common reward context): major activations by same unrewarded pictures without separation between non-reward and reward context. [From Kobayashi and Schultz (286).]

generalization induces dopamine release in nucleus accumbens (117). Fourth, the dopamine detection component increases in rewarding contexts (286), via pseudoconditioning or higher order context conditioning (367, 536, 537), without explicit reward pairing. Accordingly, reducing the separation between unrewarded and rewarded trials, and thus infecting unrewarded trials with reward context, boosts the incidence of dopamine activations to unrewarded stimuli (from 3 to 55% of neurons) **(FIGURE 10*D*)** (286), as does an increase of reward from 25 to 75% of trials (from 1 to 44% of neurons) (597, 618). Thus the initial response component reflects the detection of environmental events related to reward in the widest possible way, which conforms well with the reward function of the phasic dopamine signal.

Other neuronal systems show similar temporal evolution of stimulus responses. During visual search, frontal eye field neurons respond initially indiscriminately to target and distractor and distinguish between them only after 50–80 ms (see below, **FIGURE 38*B***) (588). Neurons in primary visual cortex V1 take 100 ms after the initial visual response to show selective orientation tuning (468) and spatial frequency selectivity (61). Over a time course of 140 ms, V1 neurons initially detect visual motion, then segregate the figure created by dot motion from background motion, and finally distinguish the relevant figure from a distractor (474). V1 and frontal eye field neurons respond initially indiscriminately to targets and distractors and distinguish between them only after 40–130 ms (434). During perceptual decision making, LIP and prefrontal neurons respond initially indiscriminately to stimulus motion and distinguish between motion strength only at 120–200 ms after stimulus onset (see below, **FIGURE 38*A***) (274, 483, 535). Pulvinar neurons show initial, nondifferential detection responses and subsequent differential stimulus ambiguity responses

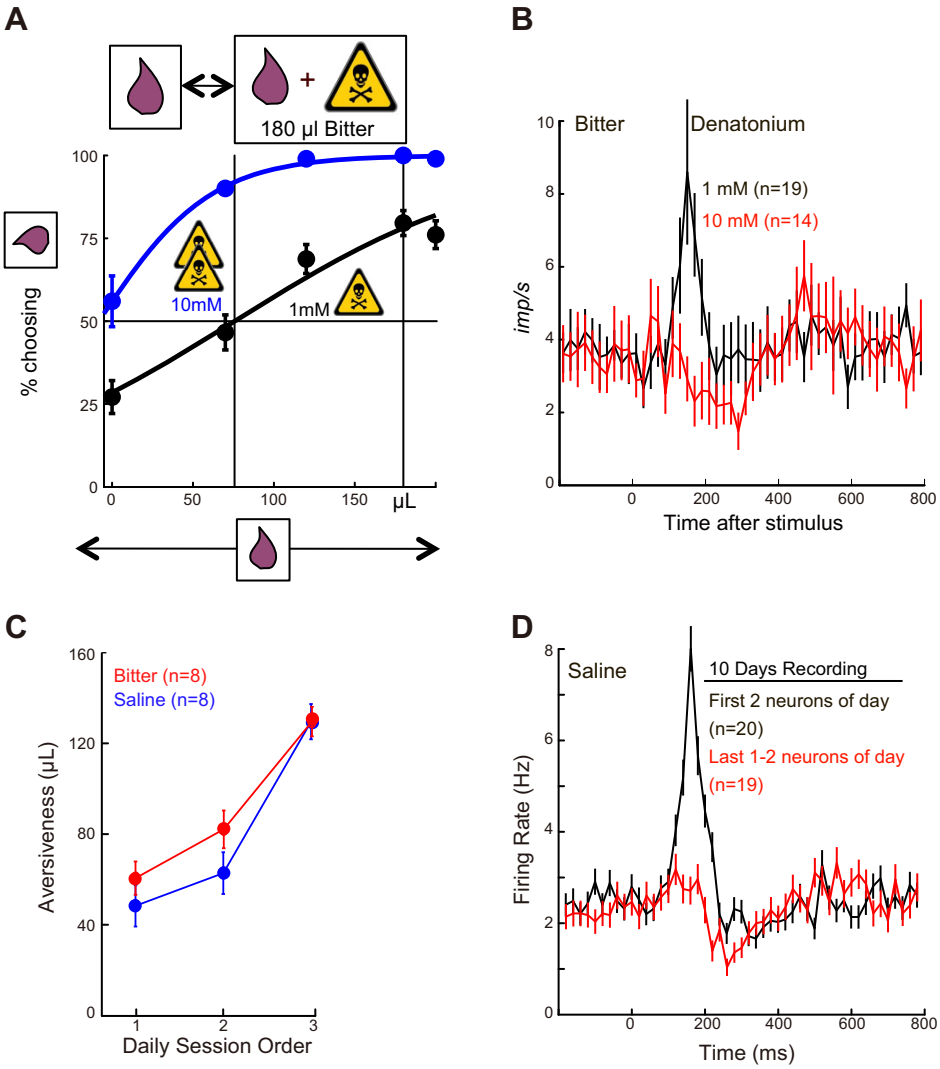


FIGURE 11. No aversive coding in monkey dopamine activations. *A:* psychophysical assessment of aversiveness of bitter solution (denatonium) as prerequisite for investigating quantitative neuronal processing of aversiveness. Monkey chose between single juice reward and juice reward + denatonium (*top*). The blue and black psychophysics curves show that increasing the volume of single juice reward induces more frequent choices of this reward against constant alternative juice + denatonium. Aversiveness of denatonium is expressed as difference between volume of single juice and juice together with denatonium at choice indifference (50% choice). Thus 1 mM denatonium is worth −100 µl of juice (black), and 10 mM denatonium is worth −180 µl (blue). The *x*-axis shows volume of single juice, and *y*-axis shows percent of choices of single juice. *B:* inverse relationship of neuronal activations to psychophysically quantified aversiveness of bitter solutions (behavioral method shown in *A*). *N* = number of neurons. Imp/s indicate firing rate. *C:* development of behavioral aversiveness of solutions within individual test days, as assessed with method shown in *A*. Liquid solutions of salt or bitter molecules lose reward value through gradual satiation and thus become increasingly aversive (thirsty monkeys work for mildly "aversive" saline solutions, and thus find them rewarding, as outcome value is dominated by liquid over salt). *D:* decreasing dopamine activations with increasing aversiveness of saline solution within individual test days (behavioral aversiveness assessment shown in *C*), suggesting subtraction of negative aversive value from gradually declining juice value due to satiation. [*A–D* are from Fiorillo et al. (160).]

during visual categorization (288). Tactile responses in mouse barrel cortex require 200 ms to become differential (300). Similar multicomponent responses are seen with reward coding. Amygdala neurons initially detect a visual stimulus and may code its identity and then transition within 60–300 ms to differential reward value coding (9, 422, 428). V1 and inferotemporal cortex responses show initial visual stimulus selectivity and only 50–90 ms later distinguish reward values (371, 557). Thus there is a sequence in the processing of external events that advances from initial detection via identification to valuation. Whereas sensory processing involves only the first two steps, reward processing requires in addition the third step.

8. No aversive dopamine activation

Aversive stimuli are well known for >30 years to activate 10–50% of dopamine neurons in awake and anesthetized monkeys, rats, and mice (102, 193, 255, 345, 362, 366, 525, 604) and to enhance dopamine concentration in nucleus accumbens (75). Five neurons close to juxtacellularly labeled ventral tegmental area (VTA) dopamine neurons were activated by footshocks (64), similar to the effects of airpuffs and footshocks on non-dopamine neurons in the area (102, 582). The activations may have derived from dopamine neurons or not; the 30–45% of dopamine neurons lacking spontaneous activity (76, 166, 176, 211) may include nonactivated neurons and thus go undetected while taking up the label. In monkeys, some putative dopamine neurons, often located above substantia nigra, are activated by air puffs (345), although they respond more frequently to conditioned stimuli than primary rewards and thus defy the higher aversiveness of primary than conditioned punishers. However, just as rewards, punishers have distinct sensory, attentional, and motivating components, and none of these studies has distinguished between them. A recent psychophysical study that made this distinction by varying these components independently reports that dopamine activations by aversive stimuli reflect the physical, sensory stimulus impact rather than their aversiveness **(FIGURE 11)** (160). Once the sensory impact is controlled for, the aversive nature of stimuli either fails entirely to affect dopamine

neurons (157), reduces the total reward value of liquid solutions and their evoked dopamine activations (157, 160), and induces outright depressant dopamine responses (73, 102, 345, 362, 366, 608). In awake animals, the aversive nature of stimuli does not generate dopamine activations, but induces negative reward prediction errors either because of the negative value or by not being rewarding, and correspondingly induces depressant dopamine responses (157). Thus the different fractions of "aversively" activated dopamine neurons seen in earlier studies likely reflect the different intensities of physical stimuli employed (159). The graded activations with punisher probability (345) might reflect salience differences of conditioned stimuli (286), or graded anticipatory relief to punisher termination (75) that is considered as rewarding (552). Salience, rather than negative reinforcer value, might also explain the surprising higher incidence of activations to conditioned stimuli compared with primary punishment (345). The sensory rather than negative value nature of dopamine responses to aversive stimuli would also explain the dopamine activations by aversive stimuli under anesthesia (64, 525, 604) where motivational impact is ruled out.

The "aversive" dopamine activation constitutes a short latency response that is shorter than the reward response, is substantially boosted by generalization from rewarded stimuli, and is often curtailed by a subsequent depression (366). With these characteristics, the aversive dopamine activations seem to constitute the initial dopamine detection response that reflects mostly physical salience. In agreement with this interpretation, the aversive dopamine neurons are also driven by rewards and unrewarded stimuli and thus are highly sensitive to a large range of stimuli (366) compatible with coding physical impact rather than motivational value. Furthermore, dopamine neurons are neither depressed nor activated by negative aversive prediction errors (157, 255, 345), and show only enhanced activations with positive aversive prediction errors possibly reflecting surprise salience. Thus the current evidence suggests that aversive stimuli activate some dopamine neurons through physical impact rather than through their negative, aversive motivational value, although some dopamine neurons with truly aversive activations can never be completely ruled out.

9. Comprehensive account of phasic dopamine response

Of the two phasic dopamine response components **(FIGURE 12)**, the initial detection activation occurs before the neurons have identified the current reward value of the stimulus and arises even with motivationally neutral events and punishers. It increases with sensory impact, novelty, reward generalization, and rewarded contexts. Thus the initial component detects potential rewards and provides an early report before having identified the value of the event properly, which relates closely to the reward coding of dopamine neurons. The second response component arises once the

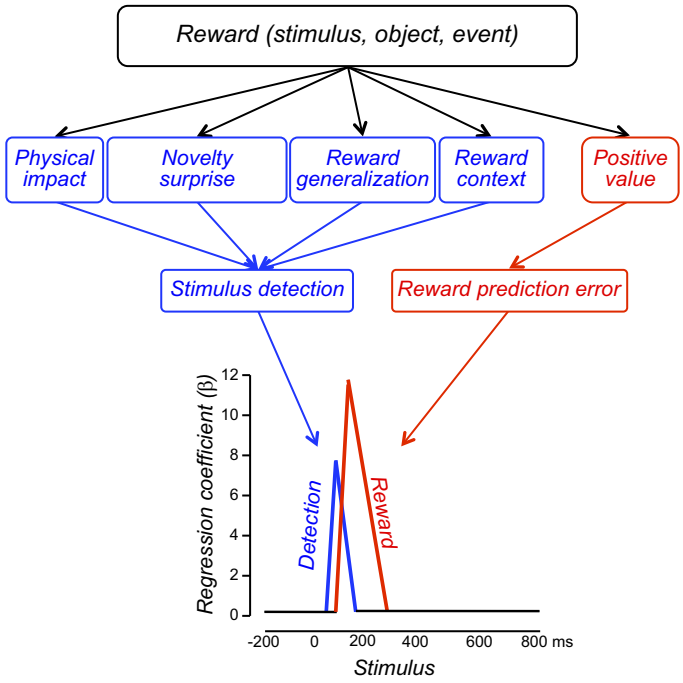


FIGURE 12. Reward components inducing the two phasic dopamine response components. The initial component (blue) detects the event before having identified its value. It increases with sensory impact (physical salience), novelty (novelty/surprise salience), generalization to rewarded stimuli, and reward context. This component is coded as temporal event prediction error (389). The second component (red) codes reward value (as reward prediction error).

neurons have well identified the reward value of the stimulus and codes only reward. As longest and strongest component it constitutes the major phasic dopamine influence on postsynaptic processing. The two components are integral parts of the dopamine reward prediction error response, which seems to be split into an early temporal error component and a subsequent value error component (also time sensitive), thus providing an interesting biological instantiation of a TD reinforcement teaching signal (574).

The phasic dopamine reward signal is homogeneous in latency, duration, and sensitivity across different dopamine neurons and very similar between different rewards, thus contrasting sharply with the widely varying reward processing in other brain structures. Graded variations in dopamine responses derive from the initial detection response component that varies in a continuous manner across the mediolateral or dorsoventral extent of the ventral midbrain. If this graded initial component occurs with aversive stimuli and is interpreted as aversive rather than detection response, some dopamine neurons may seem categorically different from those showing a smaller or no detection component (64, 255), suggesting neuronal response heterogeneity (345). However, these variations, as well as variations in the subsequent reward prediction error response (159), conform to a single-peak probability distribution rather than amounting to categorical differences and distinct dopamine subpopulations. In addition to their continuous phasic response distribution, dopamine neurons are het-

erogeneous in most other anatomical, physiological, and neurochemical respects.

The initial activation component may reflect physical, motivational, and surprise salience and thus may enhance the impact of the prediction error response in reward learning compatible with associability learning rules (425, 336). In being highly sensitive to novelty, reward generalization, and rewarded contexts, this component may reflect an early assumption about the nature of an event that is novel, resembles a reward, or occurs within a generally rewarding context. Such events have a chance to be rewards and thus are potential rewards. In responding to such stimuli, dopamine neurons detect a maximal range of potentially rewarding events and objects. Through the very short latency of these responses, dopamine neurons detect these stimuli very rapidly, even before having identified their value, and can rapidly initiate neuronal processes for approach behavior. Once the stimulus has been identified and turns out not to be a reward, it is still time to alter neuronal processing and modify or halt behavioral reactions. If the stimulus is indeed a reward, precious time may be gained and the reward approached before anybody else without such a rapid detection system arrives. By overreacting and processing also potential rewards, the mechanism would prevent premature asymptotes in reward detection and minimize reward misses (avoiding "I do not move unless the reward is for sure"). Such a mechanism would be particularly important when sparse resources present challenges for survival. With early detection facilitating rapid behavioral reactions, the two-component nature of the phasic dopamine response results in better reward acquisition and thus provides competitive evolutionary advantages.

The phasic dopamine reward signal is only one manifestation of dopamine function in the brain. In addition to the distinct and slightly less phasic dopamine risk signal (see below), dopamine exerts a tonic enabling effect on postsynaptic neurons, whose function is considerably more heterogeneous than the rather stereotyped dopamine prediction error response, even when considering its initial detection component (518). Many of these tonic functions depend on the function of the brain structure influenced by tonic extracellular dopamine concentrations, as inferred from behavioral and neuronal alterations arising from inactivations, receptor agonist or antagonist applications, and polymorphisms. These functions extend from motor processes in the striatum evidenced by Parkinsonian movement deficits to prefrontal processes of working memory, attention, reversal learning, categorization, and response control (470). An active involvement of phasic dopamine signals in these vastly different functions is questionable, in particular for functions that are recovered by dopamine receptor agonists without reinstating phasic dopamine signaling. Parkinsonism does not result from deficient dopamine reward prediction error signaling. Thus dopamine function in the brain is vast and heterogeneous and extends well beyond the phasic dopamine reward signal. Serotonin has a similar tonic enabling function, which includes aversive and reward processes (100, 201, 481, 534, 635). The question "what is dopamine doing" reflects the initial idea of "one neurotransmitter-one function" that may originate from "one brain structure-one function." In view of the many different neuronal dopamine processes, a better question might be "what is the function of the specific dopamine process we are looking at." For the phasic dopamine signal, the answer seems to be bidirectional reward prediction error coding, with additional, slightly slower risk coding.

10. Dopamine responses during reward learning

During learning episodes, dopamine responses show gradual changes that reflect the evolution of prediction errors. Dopamine neurons acquire responses to reward-predicting stimuli (322) that transfer back to the first reward-predicting stimulus in longer task schedules (144, 521). Their activations disappear gradually with extinction (598), possibly due to inhibition from neighboring pars reticulata neurons activated during extinction (408). When learning several tasks in sequence, positive reward prediction errors are generated every time a reward occurs during initial trials but subside once rewards become predictable and performance approaches asymptote. Dopamine neurons show corresponding prediction-dependent reward responses during these periods **(FIGURE 13, *A* AND *B*)** (322, 521). A temporal difference model using dopamine-like prediction errors replicates well behavioral learning and performance in these tasks **(FIGURE 13*C*)** (573). During learning set performance with one rewarded and one unrewarded option, reward is initially predicted with $P = 0.5$. Chance selection of each option leads to positive and negative 0.5 prediction errors, respectively, and to corresponding phasic dopamine activations and depressions during initial learning trials (221). The error responses decrease with increasingly correct choices.

The transfer of dopamine activation from "unconditioned" reward to reward-predicting stimuli constitutes conditioning formalized by the Rescorla-Wagner and TD learning rules. Straightforward implementations of the TD model assume gradual TD error backpropagation via imaginary "internal" stimuli (374, 524). However, biological dopamine neurons do not show such gradual backpropagating responses (162) but transfer in single steps from reward to preceding actual stimuli and show simultaneous responses to both reward and stimuli with intermediate learning steps (410). Dopamine concentration transients in ventral striatum show similar transfer **(FIGURE 14*A*)** (117, 598). Biologically plausible TD implementations and spectral timing models achieve the single step response transfer to an earlier stimulus with stimulus and eligibility traces that mark recently active

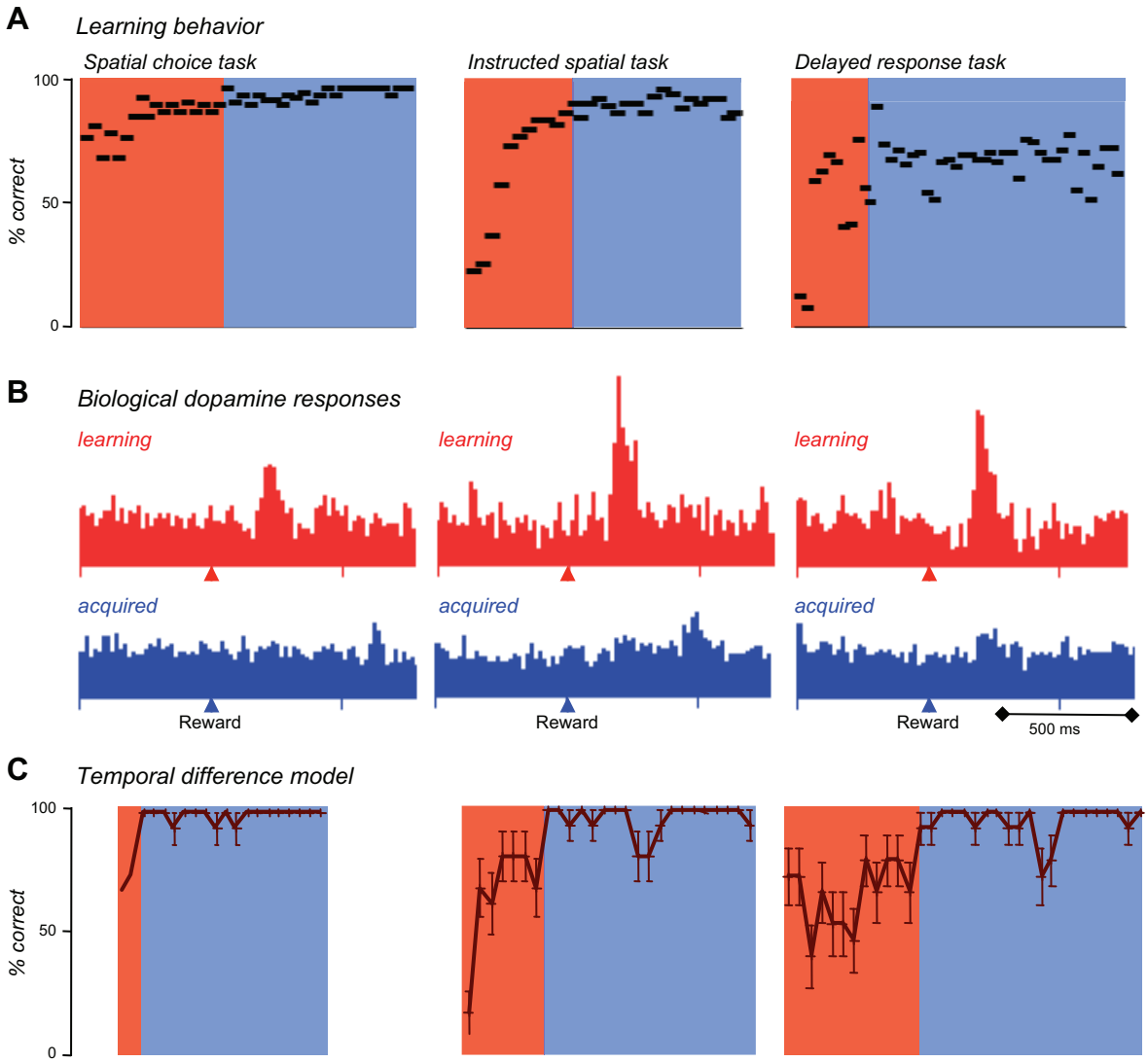


FIGURE 13. Dopamine prediction error responses reflect changing reward prediction during learning. *A:* stepwise learning of spatial delayed response task via intermediate spatial subtasks. Each dot shows mean percentage of correct behavioral performance in 20–30 trials during learning (red) and asymptotic performance (blue). *B:* positive dopamine reward prediction error responses during learning of each subtask, and their disappearance during acquired performance. Each histogram shows averaged responses from 10–35 monkey dopamine neurons recorded during the behavior shown in *A*. [*A* and *B* from Schultz et al. (521).] *C:* behavioral learning and performance of temporal difference model using the dopamine prediction error signals shown in *B*. The slower model learning compared with the animal's learning of the final delayed response task is due to a single step increase of delay from 1 to 3 s in the model, whereas the delay increased gradually in animals (*A*). [From Suri and Schultz (573), with permission from Elsevier.]

postsynaptic neurons for plasticity, as originally suggested **(FIGURE 14, *B–D*)** (71, 410, 573, 574). The eligibility traces may consist of intracellular calcium changes (154, 637), formation of calmodulin-dependent protein kinase II (236), IP_3 increase in dopamine neurons (203), and sustained neuronal activity in striatum and frontal cortex (522) reflecting previous events and choices in prefrontal neurons (30, 531). Thus use of eligibility traces results in good learning without small step backpropagation via imaginary stimuli (410, 573). These models account also better for the temporal sensitivity of prediction error responses (113, 573).

The acquisition of dopamine responses to conditioned stimuli enhances the strength of excitatory synapses onto dopamine neurons (568), and cocaine administration increases spike-time-dependent LTP in dopamine neurons (19, 318). In elucidating the mechanism, burst stimulation of dopamine neurons, mimicking dopamine reward responses, induces NMDA receptor-dependent LTP of synaptic responses in dopamine neurons (203). Compatible with the temporal characteristics of behavioral conditioning (423, 589), the burst needs to follow the synaptic stimulation by at least 0.5 or 1.0 s, whereas postsynaptic burst omission reverses LTP, and inverse timing induces long-term depres-

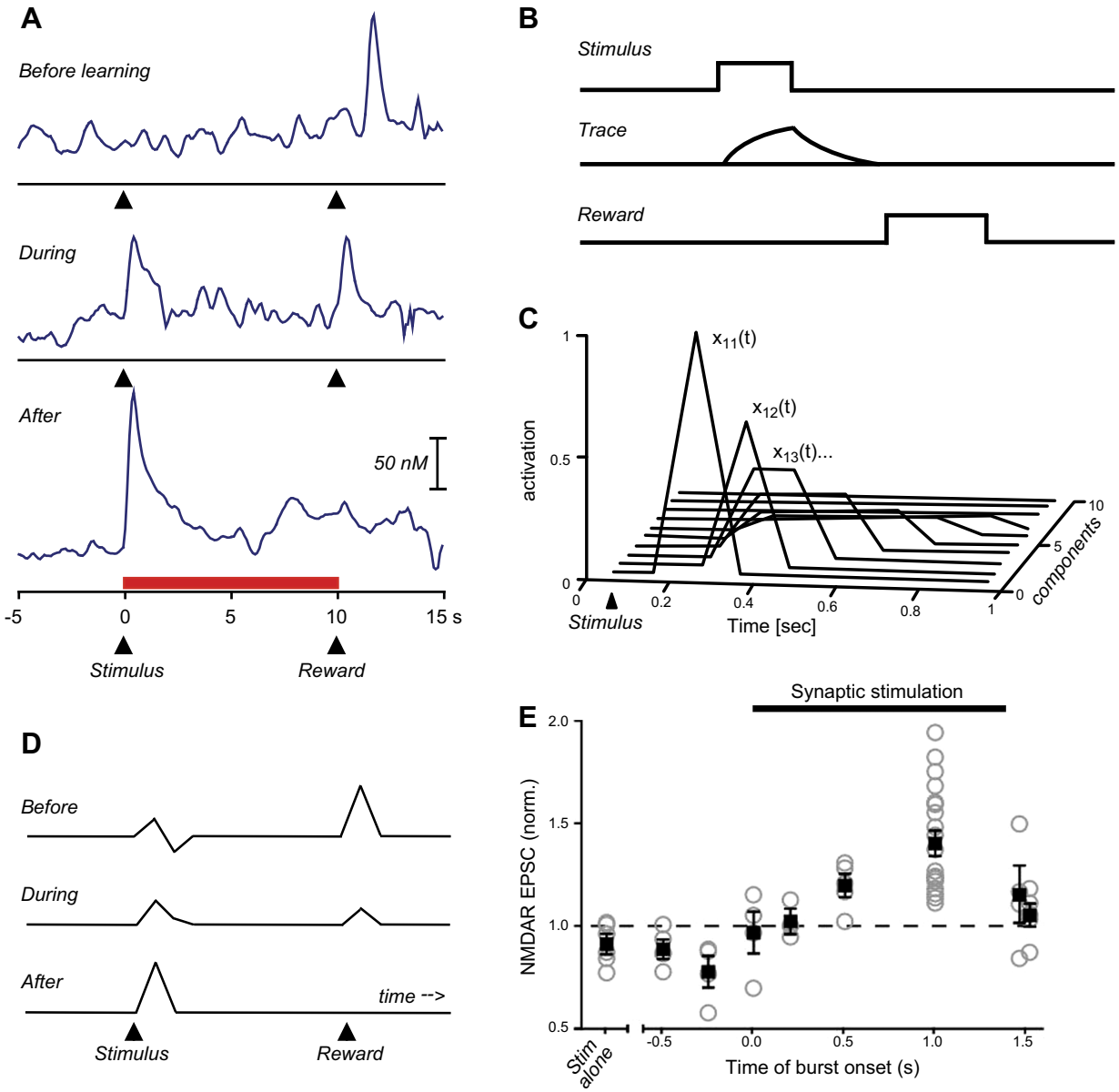


FIGURE 14. Plasticity of dopamine reward prediction error responses. *A:* transfer of dopamine response from primary reward to conditioned, reward-predicting stimulus during learning. The plots show voltammetrically measured dopamine concentration changes in ventral striatum of rats. Note simultaneous occurrence of responses to reward and stimulus during intermediate learning stage. [From Stuber et al. (598). Reprinted with permission from AAAS.] *B:* stimulus eligibility traces linking reward to stimulus during learning in original formulation of temporal difference model (TD) model. [Redrawn from Sutton and Barto (574).] *C:* stimulus eligibility traces used in biologically plausible implementation of TD model of dopamine prediction error responses. The traces allow single step transfer of dopamine prediction error signal from reward to an earlier stimulus. *D:* transfer of modeled dopamine prediction error response from reward to conditioned stimulus using eligibility traces shown in *B* and *C*. [*C* and *D* from Suri and Schultz (573), with permission from Elsevier.] *E:* time-sensitive plasticity of dopamine neurons in rat midbrain slices. LTP induction depends on timing of burst stimulation of dopamine neurons relative to their synaptic excitation. Only bursts occurring 0.5–1.5 s after synaptic stimulation lead to measurable LTP (excitatory postsynaptic currents, EPSC, mediated by NMDA receptors). [From Harnett et al. (203), with permission from Elsevier.]

sion **(FIGURE 14*E*)**. The 0.5- to 1.0-s delay is required to increase intracellular IP_3 as potential stimulus eligibility trace. Thus excitatory synapses onto dopamine neurons seem to mediate the acquisition of dopamine neurons to conditioned stimuli.

11. Non-dopamine prediction error signals

The colloquial meaning of error refers to incorrect motor performance. Climbing fibers of the cerebellum and neurons in superior colliculus and frontal and supplementary eye field

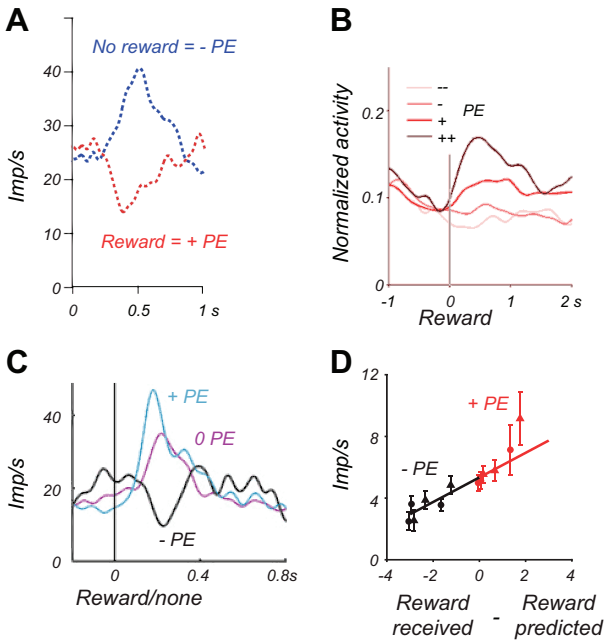


FIGURE 15. Bidirectional non-dopamine reward prediction error signals. *A:* averaged responses from 43 neurons in monkey lateral habenula during first trial of position-reward reversals. Red: positive prediction error; blue: negative prediction error. Note the inverse response polarity compared with dopamine error responses. PE = prediction error. [From Matsumoto and Hikosaka (344). Reprinted with permission from Nature Publishing Group.] *B*: averaged responses from 8 neurons in rat striatum. Subjective reward values (tiny, small, large, huge) are estimated by a Rescorla-Wagner reinforcement model fit to behavioral choices. [From Kim et al. (275).] *C:* response of single neuron in monkey amygdala. [From Belova et al. (36), with permission from Elsevier.] *D*: response of single neuron in monkey supplementary eye field. [From So and Stuphorn (551).]

signal errors of arm or eye movements (182, 266, 278, 292, 395, 569, 607). Cerebellar circuits implement a Rescorla-Wagner type prediction error for aversive conditioning (273, 357). The identification of bidirectional dopamine error signaling extended neuronal error coding to reward.

Bidirectional reward prediction error signals occur in several brain structures. Lateral habenula neurons show bidirectional reward prediction error signals that are sign inverted to dopamine responses and may affect dopamine neurons via inhibitory neurons **(FIGURE 15*A*)** (344). Select groups of phasically and tonically firing neurons in the striatum and globus pallidus code positive and negative reward prediction errors bidirectionally **(FIGURE 15*B*)** (15, 134, 227, 275, 402). Some neurons in the amygdala display separate, bidirectional error coding for reward and punishment **(FIGURE 15*C*)** (36). In the cortex, select neurons in anterior cingulate (261, 531) and supplementary eye field **(FIGURE 15*D*)** (551) code reward prediction errors bidirectionally. The bidirectional reward prediction error responses in these subcortical and cortical neurons are straightforwardly appropriate for affecting plasticity at specific postsynaptic neurons for reinforcement learning.

Positive and negative components of reward prediction errors are also coded in select groups of non-dopamine reward neurons. Some neurons in the pedunculopontine nucleus are activated by both predicted and unpredicted rewards and by conditioned stimuli (136, 396, 410) but fail to show depressions to omitted rewards and thus process reward without producing a clear error signal (281). Pedunculopontine neurons with midbrain projections above substantia nigra show bidirectional error responses to reward versus no-reward-predicting cues, but only positive prediction error responses to the reward (228). Positive responses to unpredicted but not predicted rewards occur in norepinephrine neurons as part of their attentional sensitivity (22), and in nucleus basalis Meynert neurons (466). In the striatum, some phasically or tonically firing neurons are influenced by unpredicted reward delivery or unpredicted reward omission (21, 255, 275, 556). Some of these striatal responses are stronger in Pavlovian than operant tasks (17) or occur only after particular behavioral actions (556), suggesting selectivity for the kind of behavior that resulted in the reward. In the amygdala, some neurons are more activated by unpredicted than predicted rewards (36, 40) or show unidirectional, rectified activations with both unpredicted reward delivery and omission (475). In anterior cingulate and dorsolateral prefrontal cortex, activations induced by reward omission or erroneous task performance are known for >30 years (386). In some orbitofrontal neurons, unpredicted rewards outside of tasks elicit activations, whereas reward omission fails to elicit responses (602). Some neurons in anterior and posterior cingulate and supplementary eye field are activated by unpredicted delivery or omission of reward (11, 21, 246, 261, 354, 551) and by unpredicted positive or negative feedback provided by conditioned visual reinforcers (344). Unpredictedness enhances existing reward responses in anterior cingulate cortex (206). The unidirectional activations to unpredicted delivery or omission of reward may reflect surprise salience or rectified positive and negative reward prediction errors. In reflecting surprise salience, they may code the learning rate parameter α of attentional learning rules (425) according to *Equation 3* and serve to adjust the speed of learning (475).

12. Non-dopamine responses during reward learning

Tonically active striatal interneurons (TANs) acquire discriminant responses to reward-predicting auditory stimuli within 15 min of pairing with liquid, and lose these responses within 10 min of extinction (14). Neurons in rat striatum, orbitofrontal cortex, and amygdala acquire discriminatory responses to movement-instructing and reward-predicting visual, auditory, and olfactory stimuli during learning and reversal (253, 422, 512). The acquired responses are rapidly lost during extinction (29).

Repeated learning of new stimuli leads to learning set performance in which animals acquire novel stimuli within a few trials (202), thus allowing to record from single neurons

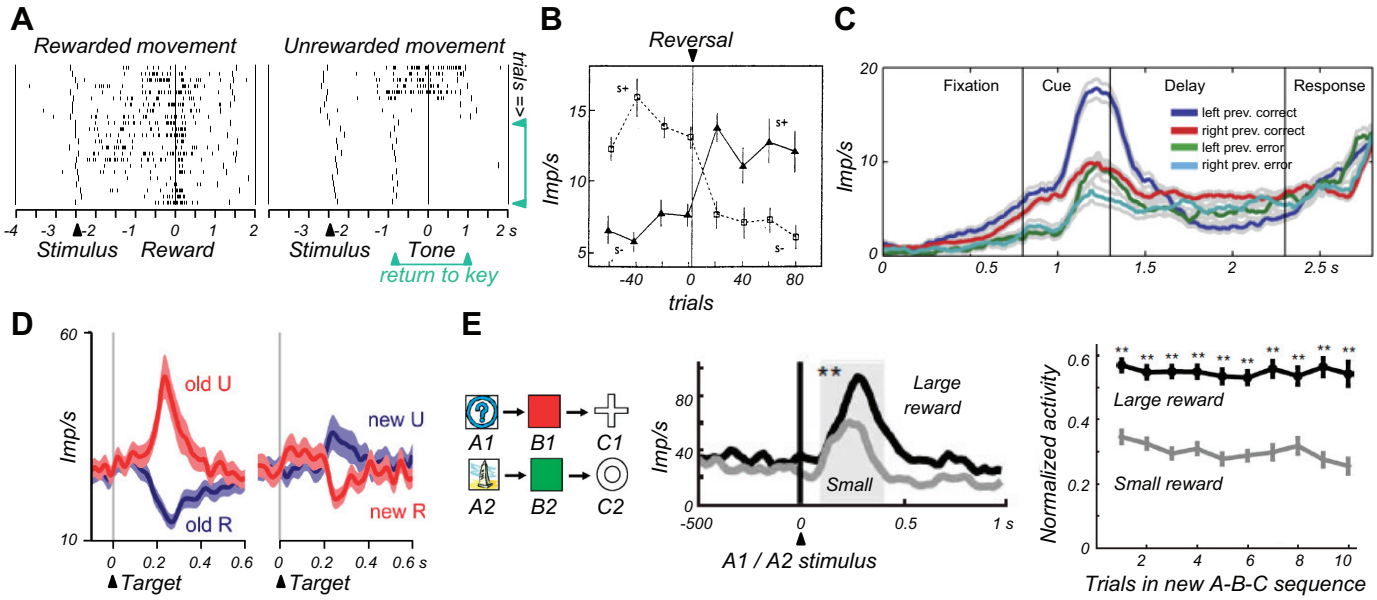


FIGURE 16. Overview of reward learning-related responses in monkey non-dopamine neurons. *A:* adaptation of reward expectation in single ventral striatum neuron during learning. In each learning episode, two new visual stimuli instruct a rewarded and an unrewarded arm movement, respectively, involving the acquisition of differential reward expectations for the same movement. Reward expectation is indicated by the return of the animal's hand to the resting key. This occurs with all rewarded movements after reward delivery (long vertical markers in left rasters, right to reward). With unrewarded movements, which alternate pseudorandomly with rewarded movement trials, the return occurs in initial trials after the tone (*top right*) but subsequently jumps before the tone (green arrows), indicating initial default reward expectation that disappears with learning. The reward expectation-related neuronal activity shows a similar development during learning (from *top* to *bottom*). [From Tremblay et al. (600).] *B*: rapid reversal of stimulus response in orbitofrontal neuron with reversed stimulus-reward association. S+ and S− are two different visual stimuli that are initially rewarded and unrewarded, respectivey. With reversal, they stay physically constant but inverse their reward prediction. [From Rolls et al. (485).] *C:* cue response in dorsolateral prefrontal cortex neuron reflecting correct, as opposed to erroneous, performance in previous trial in a delayed conditional motor task with reversal. This typical prefrontal neuron discriminates between left and right movements. Errors reduce differential left-right movement-related activity (previous trial correct: blue and red vs. error: green and blue). [From Histed et al. (217), with permission from Elsevier.] *D*: inference-related reversal of neuronal population responses in lateral habenula. These neurons are activated by unrewarded targets and depressed by rewarded targets. *Left*: in the first trial after target-reward reversal, before any new outcome occurs, the neurons continue to show activation to the old unrewarded target (red, old U) and depression to the old rewarded target (blue, old R). *Right*: in the second trial after reversal, after having experienced one outcome, the neuronal responses reflect the reward association of the other target. Thus the neurons are activated by the newly unrewarded target (blue, new U) and depressed by the newly rewarded target (red, new R), based entirely on inference from the outcome of only one target. [From Bromberg-Martin et al. (68).] *E:* inference of reward value from differentially rewarded paired associates in prefrontal cortex. Monkeys are first trained with two complete A-B-C sequences (*left*). Then, before the first trial of a new sequence, the two C stimuli are followed by large and small reward, respectively. For testing, the animal is presented with an A stimulus and choses the corresponding B and then C stimuli from the first new sequence trial on, even before sequential A-B-C-reward conditioning could occur. Neurons show corresponding reward differential activity (center, single neuron) from the first trial on (right, population average from 107 neurons), suggesting inference without explicit linking A and B stimuli to reward. [From Pan et al. (411). Reprinted with permission from Nature Publishing Group.]

during full learning episodes. Striatal and orbitofrontal neurons show three forms of reward related activity during learning (626, 600, 603, 639). Some neurons fail initially to respond to novel stimuli and acquire reward discriminatory responses during learning. Apparently they code valid reward predictions. To the opposite, other neurons respond initially indiscriminately to all novel stimuli and differentiate as learning advances, possibly reflecting exploration. A third group of neurons shows reward expectation activity that reflects an initial general expectation and becomes selective for rewarded trials as learning progresses **(FIGURE 16*A*)**. The increases in learning rate with electrical striatal stimulation suggest that some of these striatal neurons may mediate reward learning (639).

In analogy to learning set, repeated reversals of reward predictions of the same stimuli lead to reversal set performance in which animals switch after a single reversed trial

to the alternative option rather than requiring multiple learning trials, even without having experienced a reward. Prefrontal, orbitofrontal, and amygdala neurons show rapid response loss to previously learned stimuli and acquisition of the currently rewarded stimuli (626), closely corresponding to behavioral changes **(FIGURE 16*B*)** (422, 485). The changes occur several trials earlier in striatal neurons compared with prefrontal neurons (421), but are similar between orbitofrontal and amygdala neurons (378). Some striatal and cortical responses reflect the correct or incorrect performance of the previous trial in reversal sets, thus bridging information across consecutive trials **(FIGURE 16*C*)** (217).

Both learning sets and reversal sets define task structures and behavioral rules which allow individuals to infer reward values of stimuli without having actually experienced the new values. During reversals, responses in globus pallidus, lateral habenula, and dopamine neurons reflect the new reward value of a target already after the first reversed trial, based solely on inference from the other target **(FIGURE 16*D*)** (68). Neuronal responses in dorsolateral prefrontal cortex reflect reward value based on inference from paired associates **(FIGURE 16*E*)** (411) or transitivity (409). Neuronal responses in the striatum reflect inference by exclusion of alternative stimuli (409). These responses seem to incorporate the acquired rule into their responses, supposedly by accessing model-based learning mechanisms.

C. Dopamine Implementation of Reward Learning

1. Origin of reward prediction error response

The dopamine reward prediction error signal may derive from two basic mechanisms. The dopamine neurons may receive a complete bidirectional reward prediction error signal from presynaptic input. Alternatively, they may compute the signal from heterogeneous components by subtracting the predicted from the experienced reward at the time of the reward and, in the spirit of TD models, at the time of reward predictors. The computation would require separate inputs for positive and negative errors, both at the time of the final reward and at the time of cues predicting all future rewards.

Dopamine neurons receive main reward information from striatum, amygdala, subthalamic nucleus, pedunculopontine nucleus, rostromedial reticular nucleus, and GABAergic neurons of pars reticulata of substantia nigra (625). They receive homeostatic signals from lateral hypothalamus (464). However, dorsal, medial, and orbital prefrontal cortical areas in monkeys provide probably less direct inputs to dopamine neurons than often assumed (625) but could influence them transsynaptically via the striatum.

Bidirectional reward prediction error response components may arise from lateral habenula signals **(FIGURE 15*A*)** (344) which themselves derive from globus pallidus (227) but are sign inverted compared with dopamine responses. The habenula exerts strong and exclusively inhibitory influences on dopamine neurons (95, 250, 344) via GABAergic neurons in the rostromedial reticular nucleus (226, 249). The same habenula neurons show bidirectional punishment prediction error signals opposite to their reward prediction error responses. They respond to positive reward prediction errors in a very similar way as to negative punishment prediction errors (by depression), and to negative reward prediction errors similarly as to positive punishment prediction errors (by activation), thus coding value monotonically across rewards and punishers (346). However, it is unclear why the aversive habenula input would not impact on dopamine neurons (157, 160). Bidirectional reward prediction error inputs may also arise from select neurons in a number of other structures, including the striatum **(FIGURE 15*B*)** (15, 134, 275), amygdala **(FIGURE 15C)** (36), anterior cingulate (261, 531), and supplementary eye field **(FIGURE 15*D*)** (551). To make these neurons input candidates for the bidirectional dopamine prediction error signal would require demonstration of their projection to dopamine neurons, which may become technically possible by using selective optogenetic stimulation of error processing neurons that project to dopamine neurons.

Positive dopamine prediction error response components may arise from direct excitatory inputs from pedunculopontine neurons (131) responding to sensory stimuli, reward-predicting cues, and rewards (136, 228, 396, 410). In support of this possibility, inactivation of pedunculopontine neurons reduces dopamine stimulus responses (410). Norepinephrine neurons activated by attentional reward components (22) and nucleus basalis Meynert neurons activated by rewards (466) may also project to dopamine neurons. Different groups of striatal neurons exert activating influences on dopamine neurons via double inhibition (407) and respond to reward-predicting stimuli irrespective of the stimuli being themselves predicted (222) or incorporate reward predictions (17, 21, 255, 275, 402, 556). Positive response components may also derive from amygdala neurons that are activated by rewards irrespective of prediction (422), by unpredicted rewards (40), or by unpredicted reward delivery and omission (475). Reward responses in monkey frontal cortex (11, 21, 206, 246, 261, 347, 354) may reach dopamine neurons via the striatum.

Negative dopamine error responses may arise from GABAergic inputs from VTA, substantia nigra pars reticulata, or striatum (625). Some of these inputs are activated by aversive stimuli (10, 102, 582). Optogenetic activation of pars reticulata GABAergic neurons reduces dopamine impulse activity (613). Inhibition from GABAergic pars re-

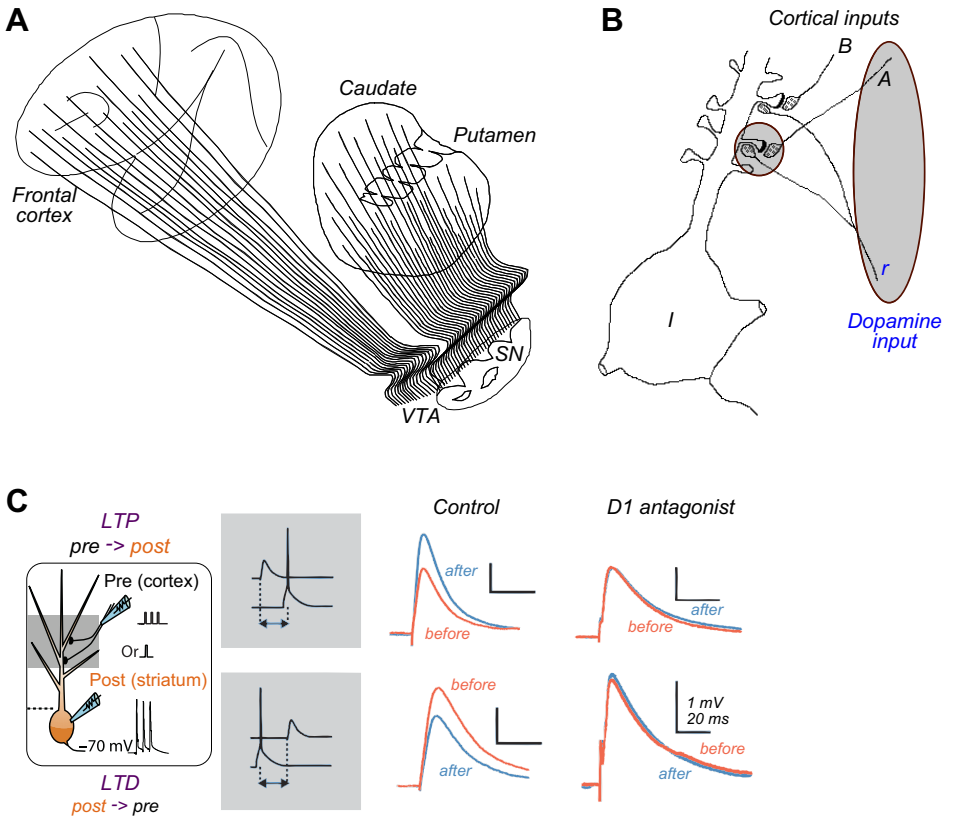


FIGURE 17. Anatomical and cellular dopamine influences on neuronal plasticity. *A:* global dopamine signal advancing to striatum and cortex. The population response of the majority of substantia nigra (SN) pars compacta and ventral tegmental area (VTA) dopamine neurons can be schematized as a synchronous, parallel volley of activity advancing at a velocity of 1–2 m/s (525) along the diverging projections from the midbrain to large populations of striatum (caudate and putamen) and cortex. [From Schultz (517).] *B:* differential influence of global dopamine signal on selectively active corticostriatal neurotransmission. The dopamine reinforcement signal (r) modifies conjointly active Hebbian synapses from active input (*A*) at striatal neuron (I) but leaves inactive synapses from inactive input (*B*) unchanged. Gray circle and ellipse indicate simultaneously active elements. There are ~10,000 cortical terminals and 1,000 dopamine varicosities on each striatal neuron (138, 192). Drawing based on data from Freund et al. (171) and Smith and Bolam (548). [From Schultz (520).] *C:* dopamine-dependent neuronal plasticity in striatal neurons. *Left*: experimental in vitro arrangement of cortical input stimulation and striatal neuron recording using a spike time dependent plasticity (STDP) protocol in rats. Control: positive STDP timing (striatal EPSP preceding action potential, Δt = 20 ms) results in long-term potentiation (LTP) (*top*), whereas negative STDP timing (striatal EPSP following action potential, Δt = −30 ms) results in long-term depression (LTD) (1, orange, and 2, blue, refer to before and after stimulation). Dopamine D1 recepor antagonist SCH23390 (10 μM) blocks both forms of plasticity, whereas D2 antagonist sulpiride (10 μM) has less clear effects and affects only plasticity onset times (not shown). [From Shen et al. (540). Reprinted with permission from AAAS. From Pawlak and Kerr (424).]

ticulata neurons showing sustained activations during reward prediction may also result in cancellation of reward activations in dopamine neurons by predicted rewards (102).

The computation of reward prediction error requires a reward prediction to be present at the time of the primary or higher order reward. This prediction may be mediated by the well-known sustained activations preceding rewards and reward-predicting stimuli in neurons of structures projecting mono- or polysynaptically to dopamine neurons, including orbitofrontal cortex (544, 602, 628), dorsal and ventral striatum (18, 205, 215, 523), and amygdala (40) (see below, **FIGURE 37, *D* AND *E***).

2. Postsynaptic influences

The phasic, time specific, reward prediction error reporting dopamine signal is propagated from specific dopamine populations in VTA and substantia nigra pars compacta (SNpc) to nucleus accumbens, dorsal striatum, frontal cortex, amygdala, and other forebrain structures. The widespread and diverging anatomical dopamine projections to postsynaptic structures show limited topography **(FIGURE 17*A*)** (177, 251, 342). Dopamine release and postsynaptic influence vary within small regions and are locally finely regulated (92). The tonic local dopamine concentration has separable, powerful enabling effects on a large variety of motor, cognitive, and motivational processes that are chal-

lenged by impaired dopamine function induced by Parkinson's disease, lesions, and psychopharmacological interventions (518). Distinction of the tonic function against the phasic signal is crucial for adequate conceptualization of dopamine's role in behavior. Importantly, interference of dopamine neurotransmission does not allow distinction between the two functions without special procedures. The deficits derive often from impaired tonic dopamine function and thus do not match the normal functions conveyed by phasic signals.

Dopamine synapses on dendritic spines of postsynaptic neurons in striatum and frontal cortex show a triad arrangement with cortical afferents that constitutes an architecture for three-factor Hebbian plasticity and allows the global dopamine reinforcement signal to exert differential influences on selectively active corticostriatal neurotransmission **(FIGURE 17*B*)** (171, 186, 461, 517). Indeed, electrical midbrain stimulation induces dopamine D1 receptor dependent long-term potentiation in striatal neurons in vivo (461) and expansion of cortical auditory fields for costimulated frequencies (28). Dopamine plays crucial roles in long-term potentiation (LTP) and depression (LTD) in striatum (82, 265, 294, 583), frontal cortex (194, 400), hippocampus (401), and amygdala (495). Dopamine turns tetanus stimulation-induced cortical LTD into LTP when paired with NMDA receptor activation (343). In spike-time-dependent plasticity (STDP) protocols, LTP occurs differentially when presynaptic stimulation precedes postsynaptic stimulation by a few tens of milliseconds, whereas LTD occurs with reverse sequence (48). Slice STDP protocols induce anatomical enlargement of striatal dendritic spines by burst-stimulating dopamine axons 0.3–2.0 s after depolarization of medium spiny striatal neurons, involving dopamine D1, but not D2, receptors and striatal NMDA receptors (642). The specific sequence of striatal excitation followed, but not preceded, by dopamine stimulation corresponds well to the temporal requirements for behavioral conditioning (423, 589). Correspondingly, dopamine D1 receptors are necessary for both LTP and LTD in striatal neurons, whereas D2 receptors may be more involved in the time course of plasticity **(FIGURE 17*C*)** (424). The effects may differ between striatal neuron types (540), as LTP in indirect pathway neurons does not depend on dopamine (294). In cultured hippocampus neurons, stimulation of dopamine D1 receptors enhances STDP LTP and turns LTD into LTP (648). These in vitro STDP data suggest a necessary role of dopamine in plasticity and should be carried to in vivo preparations for better functional assessment of their role in natural learning situations. Formal modeling demonstrates the suitability and power of dopamine-like reward prediction error signals in STDP learning (170, 247, 446), including prevention of run-away synaptic weights (375, 457).

3. Necessary dopamine involvement in learning (inactivation)

Hundreds of lesioning and psychopharmacological studies over the past 30 years demonstrate behavioral learning deficits with impaired dopamine function in a large variety of tasks. Following the three-factor Hebbian learning scheme (461, 517), intra-accumbens or systemically applied D1 receptor blockers or knockout of NMDA receptors on D1 receptor expressing striatal neurons impairs simple types of stimulus-reward learning **(FIGURE 18*A*)** (130, 163, 413). Similarly, intracortical D1 antagonist application impairs acquisition of differential neuronal responses in monkey prefrontal cortex and parallel behavioral learning in a delayed conditional motor task (440). Learning is somewhat less impaired by systemic D1 receptor blockade in tasks involving less direct reactions to stimuli and engaging less phasic dopamine responses (sign trackers versus goal trackers, **FIGURE 18*B***) (45, 163, 415). The learning deficits occur irrespective of performance deficits (163, 440, 624, 649, 650). Thus prediction error responses of dopamine neurons are not indiscriminately involved in all learning forms and

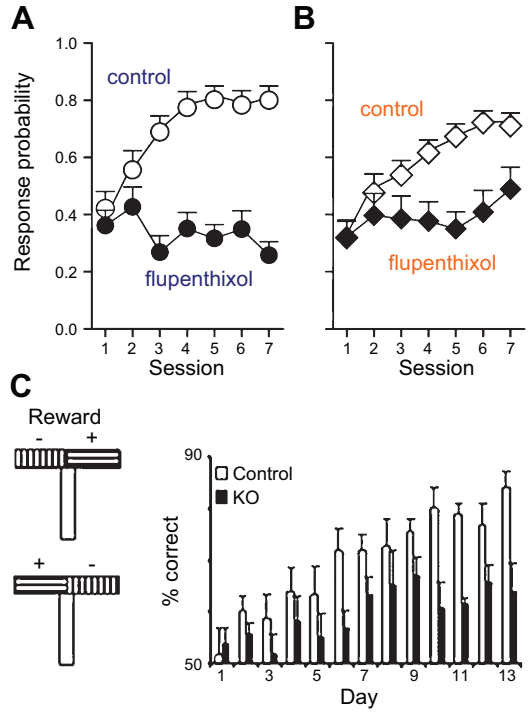


FIGURE 18. Differential learning impairment by dopamine receptor blockade. *A:* dopamine receptor blockade induces severe learning deficits in sign-tracking rats (learning under flupenthixol, not shown; test drug free, shown here). Sign trackers contact a conditioned stimulus before approaching the goal. *B:* dopamine receptor blockade by flupenthixol induces less severe learning deficits in goal-tracking rats. Goal trackers bypass the stimulus and directly approach the goal. [*A* and *B* from Flagel et al. (163). Reprinted with permission from Nature Publishing Group.] *C:* knockout (KO) of NMDA receptors on mouse midbrain dopamine neurons reduces phasic impulse bursts (not shown) and induces learning deficits in pseudorandomly alternating T-maze reward arms. [From Zweifel et al. (650).]

play more important roles when prediction errors have more direct effects.

Genetic NMDA receptor knockout on mice dopamine neurons reduces neuronal burst responses to reward-predicting stimuli and, in parallel, induce deficits in a wide range of learning situations including conditioned place preference, maze learning, and operant conditioning **(FIGURE 18*C*)** (624, 649, 650). Correspondingly, $GABA_A$ receptor knockout on dopamine neurons enhances dopamine release and T-maze and lever press learning (414). Without permitting adaptation during ontogenetic development, inactivation of dopamine neurons by local muscimol impairs learning from unexpected reward increase or omission (478). Transient direct or transsynaptic optogenetic inhibition of dopamine neurons induces place dispreference learning in choices between two compartments (see **FIGURE 19*E*** below) (244, 582).

Taken together, learning depends on intact dopamine function in simple reward contiguity situations with explicit, easily identifiable rewards and conditioned stimuli that engage phasic dopamine responses. Learning may not depend on dopamine neurons in situations in which they do not show phasic responses. Learning may also occur despite dopamine impairments when other learning systems compensate within the employed time frames, which may consist of switching to other struc-

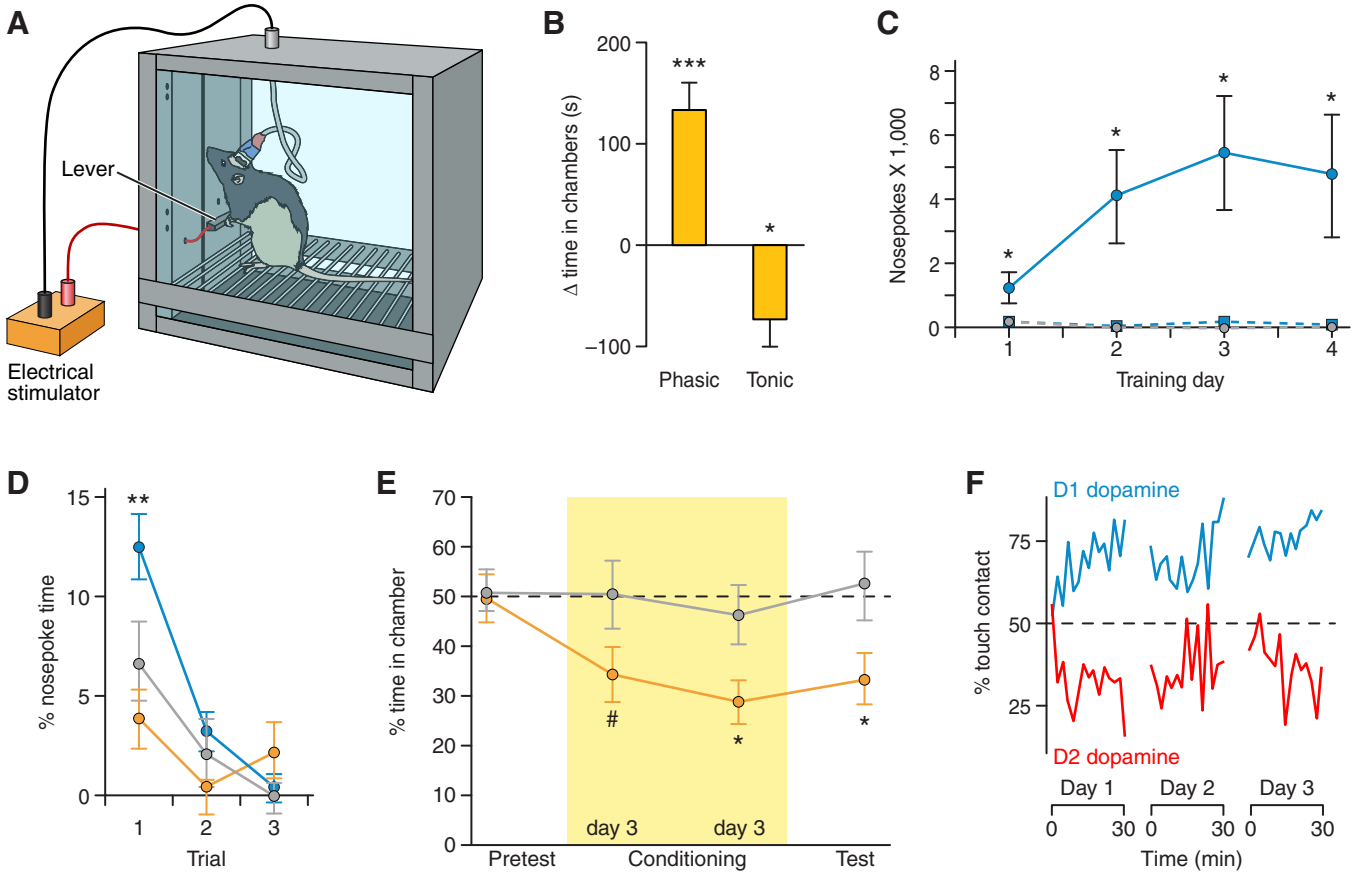


FIGURE 19. Effects of dopamine stimulation on behavioral learning. *A:* operant self-stimulation chamber. Pressing the lever elicits an electrical or optogenetic stimulus delivered to a specific brain structure through an implanted electrode/optrode. [From Rentato M. E. Sabatini, The history of electrical stimulation of the brain, available at http://www.cerebromente.org.br/n18/history/stimulation_i.htm.] *B:* place preference conditioning by phasic but not tonic optogenetic activation of dopamine neurons in mice. [From Tsai et al. (605). Reprinted with permission from AAAS.] *C:* operant nosepoke learning induced by optogenetic activation of dopamine neurons at stimulated target (blue) but not at inactive target (black) in rats. [From Witten et al. (641), with permission from Elsevier.] *D:* optogenetic activation of dopamine neurons unblocks learning of visual stimulus for nosepoke in rats, as shown by stronger response to stimulus paired with optogenetic stimulation in a blocking procedure (blue) compared with unpaired stimulus (orange) and stimulation in wild-type animals (gray). Response decrements are typical for unreinforced tests ("in extinction"). [From Steinberg et al. (562). Reprinted with permission from Nature Publishing Group.] *E:* place dispreference conditioning by direct optogenetic inhibition of dopamine neurons in mice (yellow; optical stimulation alone, gray). [From Tan et al. (582), with permission from Elsevier.] *F:* operant conditioning of approach (blue) and avoidance (red) behavior by optogenetic activation of D1 and D2 dopamine receptor expressing striatal neurons, respectively, in mice. [From Kravitz et al. (293). Reprinted by permission of Nature Publishing Group.]

tures after receptor blockade or lesions or altering ontogenetic development in gene knockouts. Fine-grained behavioral analysis may nevertheless reveal subtle remaining learning deficits. Given the crucial importance of reward for survival, multiple, flexible systems that sustain basic forms of learning despite partial impairments are biologically plausible and would enhance evolutionary fitness.

4. Sufficient dopamine involvement in learning (stimulation)

Intracranial electrical self-stimulation (398) induces operant learning of lever pressing **(FIGURE 19*A*)**. The effect is partly based on activation of VTA and SNpc dopamine neurons (103, 155) and involves dopamine release onto D1 receptors in nucleus accumbens (117). Optogenetic activation of midbrain dopamine neurons in rodents elicits learning of place preference, nose poking and lever pressing, and restores learning in a blocking procedure **(FIGURE 19, *B–D*)** (2, 244, 271, 562, 605, 641). Optogenetically induced nose poke learning occurs also with activation of dopamine axons in rat nucleus accumbens and is attenuated by intra-accumbens infusion of dopamine D1 and D2 receptor antagonists, suggesting involvement of dopamine projections to nucleus accumbens (562). The stimulations are similar in frequency to natural dopamine activations (20–50 Hz) but often exceed their natural durations of 150–250 ms by 2–7 times (500-1,000 ms), although stimulation durations of 200 ms are also effective (271, 641). The longer-than-natural activations induce more dopamine release in vitro (641) but may be compromised by the limited capacity of dopamine neurons for prolonged discharges due to pronounced tendency for depolarization block (54, 463). In contrast to 50-Hz phasic optogenetic dopamine activation, tonic (1 Hz) stimulation does not induce place preference learning (605), emphasizing the efficacy of phasic activations. In monkeys, electrical 200 Hz/200 ms VTA microstimulation induces choice preferences and neuromagnetic striatal activations, presumably reflecting dopamine activations (20). Opposite to excitations of dopamine neurons, direct optogenetic inhibition of dopamine neurons, or their indirect inhibition via activation of local, presynaptic GABA neurons, leads to place dispreference learning **(FIGURE 19*E*)** (244, 582). Conceivably the optogenetic excitation and inhibition of dopamine neurons mimic positive and negative dopamine prediction error signals and affect learning accordingly.

Learning induced by stimulation of dopamine neurons and their postsynaptic targets demonstrates the acquisition of associations with stimulation reward rather than simply reflecting arousal, response bias, direct effects on behavior, or performance enhancement, as the learned behavior is retained over several days, declines only gradually during extinction (271, 293, 582, 605) and is sensitive to devaluation by contingency degradation (641). The learning may involve striatal dopamine receptors, as optogenetic stimulation of D1 receptor expressing striatal neurons induces approach learning and stimulation of D2 receptor containing striatal neurons induces avoidance learning **(FIGURE 19*F*)** (293). The dichotomous striatal functions go back to learning models showing choice preference mediated by D1, and dispreference by D2, receptor containing striatal neurons and their respectively associated direct and indirect pathways (169). Taken together, electrical and optogenetic activation of dopamine neurons seem to mimic the natural dopamine prediction error responses to rewards and elicit learning behavior. The wider question is, to which extent would dopamine responses to a natural reward bias Pavlovian associations with that reward or operantly condition actions leading to the reward?

IV. APPROACH AND CHOICE

Rewards induce approach behavior and serve as arguments for economic choices. These functions involve a fundamental requirement, the prediction of reward that elicits reward expectation. We can approach an object because we expect it to be a reward. Without that expectation, we would approach an unknown object only when we are exploring the world. The same holds for choices. Final outcomes are often not apparent at the time of choices. Without such information, choices would be guesses and thus be very inefficient. One simply cannot make informed decisions without knowing what to expect from each option. Explorations are not informed choices but are helpful for detecting potentially better rewards. Thusgoal-directed approach behavior and informed choices are based on predictions of reward.

Predictions are acquired in the most basic form by Pavlovian conditioning. The sound of the bell predicts the sausage to Pavlov's dog. Also, the stimuli occurring during operant conditioning become Pavlovian conditioned reward predictors as the improving behavior results in more rewards. Each stimulus may become a higher order reward through conditioning and learning and thus has distinct economic value. Money does not have homeostatic or reproductive functions on its own but allows individuals to acquire nutritional and mating rewards and thus is a non-primary reward. More complex forms of prediction learning involve observation of other individuals, inference of events based on acquired knowledge, as in Bayesian updating in which the "posterior" probability of event occurrence arises from the "prior" probability, and conditional and reflective reasoning ("if I do this I can expect to obtain that reward"). In mediating Pavlovian and operant conditioning, in addition to stimulus-reward and action-reward pairing, the three-term contingency under which a discriminative stimulus signals an effective action-reward relationship constitutes the key factor for goal-directed approach behavior and informed choices. Contingency requires prediction errors, which serve to acquire and update the crucial predictions. Reward neurons are sensitive to contingency **(FIGURE 5)** (41) and code reward prediction errors **(FIGURE 7)** (517).

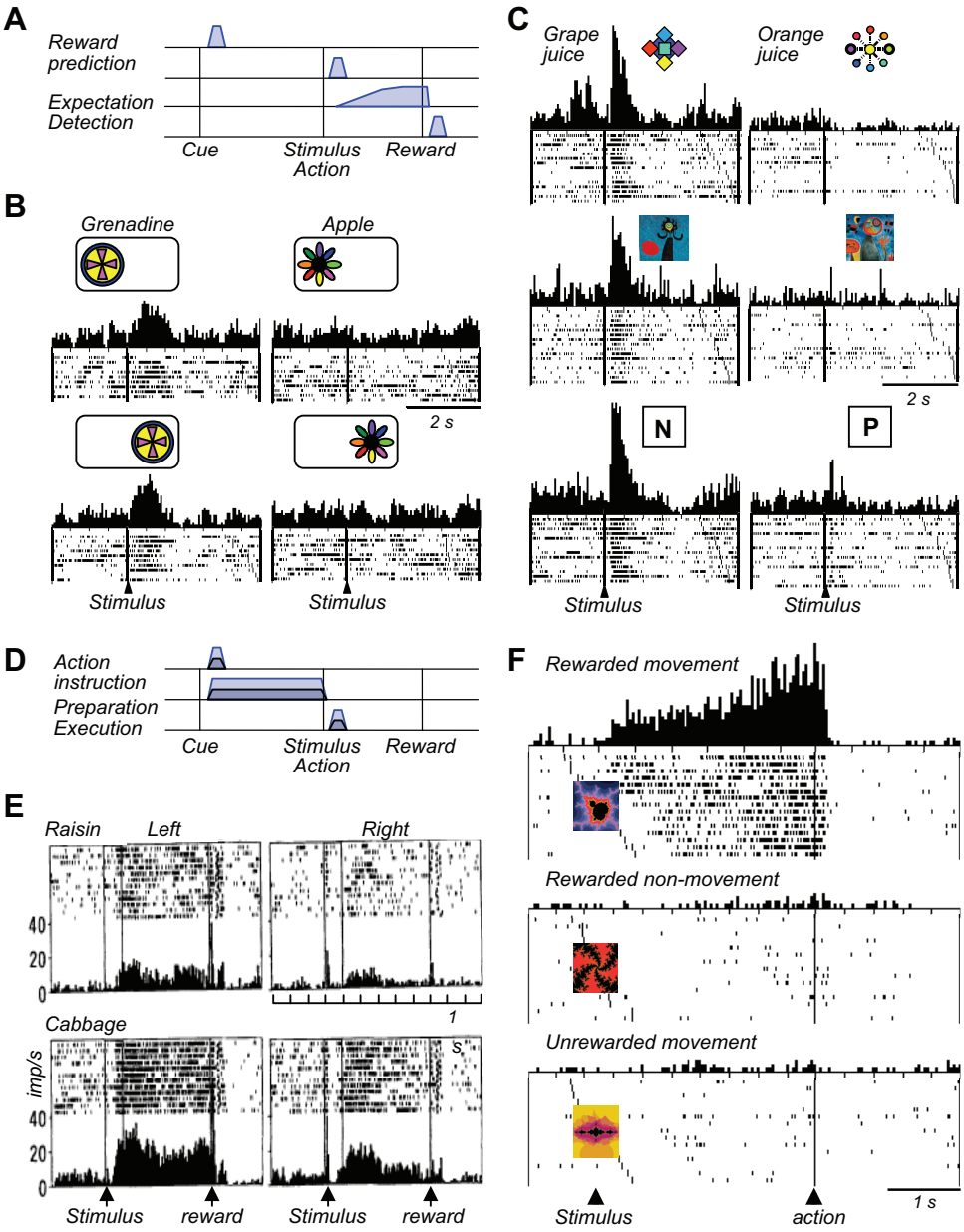


FIGURE 20. Explicit and conjoint neuronal reward signals in monkeys. *A:* scheme of explicit reward processing. Reward-predicting responses occur to an initial cue or to an action-inducing stimulus. Anticipatory activity occurs during the expectation of reward elicited by an external stimulus or action. Reward detection responses occur to the final reward. These different activities occur in separate neurons, except for dopamine neurons which all show similar responses (and no sustained reward expectation activations). *B:* neuronal reward processing irrespective of spatial position in orbitofrontal cortex. Differential activation by conditioned stimulus predicting grenadine juice but not apple juice, irrespective of spatial stimulus position and required movement (spatial delayed response task). *C:* neuronal reward processing irrespective of visual stimulus features in orbitofrontal cortex. Differential activations by conditioned stimuli predicting grape juice but not orange juice, irrespective of visual stimulus features. [*A* and *B* from Tremblay and Schultz (601).] *D:* scheme of conjoint reward-action processing. Predicted reward affects neuronal activity differentiating between different movement parameters during the instruction, preparation, and execution of action (e.g., spatial or go-nogo, blue vs. gray). *E:* conjoint processing of reward type (raisin vs. cabbage) and differentiating between spatial target positions in dorsolateral prefrontal cortex (spatial delayed response task). [From Watanabe (627). Reprinted with permission from Nature Publishing Group.] *F:* conjoint processing of reward (vs. no reward) and movement (vs. no movement) in caudate nucleus (delayed go-nogo task). The neuronal activities in *E* and *F* reflect the specific future reward together with the specific action required to obtain that reward. [From Hollerman et al. (222).]

A. Basic Reward Processing

1. Events eliciting reward responses

Explicit neuronal reward signals code only reward information. They occur as phasic responses to the delivery of "unconditioned" rewards, as phasic responses to conditioned, reward-predicting stimuli, and as sustained activity during the expectation of reward **(FIGURE 20A)**. They reflect only reward properties and do not vary with sensory or motor aspects **(FIGURE 20, *B* AND *C*)**.

"Unconditioned" liquid or food rewards elicit explicit reward signals in all main components of the brain's reward system (see **FIGURES 7**; **8, *A–C***; **13*B***; **AND 27*A***) (16, 18, 42, 60, 205, 215, 322, 383, 387, 405, 422, 451, 492, 541, 592). Reward responses are found also in premotor, prefrontal, cingulate, insular, and perirhinal cortex (8, 246, 369, 569). Some prefrontal neurons code the reward of the preceding trial (30, 179).

Conditioned, reward-predicting stimuli induce phasic explicit reward signals in all main reward structures of the brain, including orbitofrontal cortex, striatum, amygdala, and dopamine neurons (VTA and SNpc) (**FIGURE 20, *B* AND *C***; see also **FIGURES 5*B***; **7*C***; **8*C***; **9, *A*, *B*, AND *D***; **27, *B* AND *D***) (e.g., Refs. 41, 205, 276, 319, 371, 387, 405, 422, 516, 542, 601). These responses occur irrespective of the physical properties of the conditioned stimuli. Similar reward-related stimulus responses occur also in dorsolateral prefrontal cortex and anterior insula (338, 369, 411).

During reward expectations evoked by reward-predicting stimuli, select neurons in orbitofrontal cortex, striatum, and amygdala show slower, sustained explicit reward signals (see **FIGURES 16*A* AND 38, *D* AND *E***) (18, 37, 215, 601) which may also reflect reward timing (528) and temporal reward structure (40).

In addition to inducing explicit reward signals, rewards affect existing sensory-specific and action-specific activity. Neurons coding reward together with differential sensory information are found in orbitofrontal, dorsolateral prefrontal, perirhinal, and inferotemporal cortex (231, 319, 371, 394, 411). Reward affects neuronal activity differentiating between different movement parameters during the instruction, preparation and execution of action in prefrontal and premotor cortex **(FIGURE 20, *D* AND *E*)** (135, 282, 313, 348, 476, 606, 627), anterior and posterior cingulate cortex (354, 542), parietal cortex (381, 427, 433), striatum **(FIGURE 20*F*)** (107, 205, 222, 260, 308), globus pallidus (178), substantia nigra pars reticulata (502), superior colliculus (243), and amygdala (428). By processing information about the forthcoming reward during the preparation or execution of action, these activities may reflect a representation of the reward before and during the movement toward the reward, which fulfills a crucial requirement for goal-directed behavior (133). However, in motor structures, increased movement-related activity with larger rewards may reflect the more energized movements rather than representing a true conjoint reward-action signal (476). After reward delivery, responses in dorsolateral prefrontal neurons differentiate between movement directions (606).

Taken together, reward neurons show passive responses to reward-predicting stimuli and rewards, and sustained activities in advance of predicted rewards. Explicit reward signals in a limited number of brain structures reflect reward information but code neither sensory nor motor information. In contrast, sensory and motor neurons in several brain structures process reward information conjointly with sensory responses and action activity **(FIGURE 21)**. Reward coding involves the large majority of dopamine neurons, which all respond in a very similar manner to the effective events. In contrast, only specific and much smaller fractions of neurons in the other reward structures show reward activity, and their responses can be quite selective for one or the other of several task events.

2. Subjective reward perception

Neuronal reward responses may depend on the subjective perception of reward-predicting stimuli. In a signal detection task, monkeys correctly or incorrectly report the presence of stimuli (hit and false alarm, respectively), and correctly or incorrectly report stimulus absence (correct rejection and miss) (122). Dopamine neurons are activated by an initial task stimulus only when the animal correctly reports its detection, whereas they are not activated by the same physical stimulus when the animal reports its absence (miss). Thus subjective perception of the stimulus is necessary to activate dopamine neurons. However, subjective perception is not sufficient, as dopamine neurons do not respond when the physically absent stimulus is incorrectly reported as present (false alarm) **(FIGURE 22)**. The prediction error responses to a subsequent cue and to the final reward reflect also the subjective perception of the initial

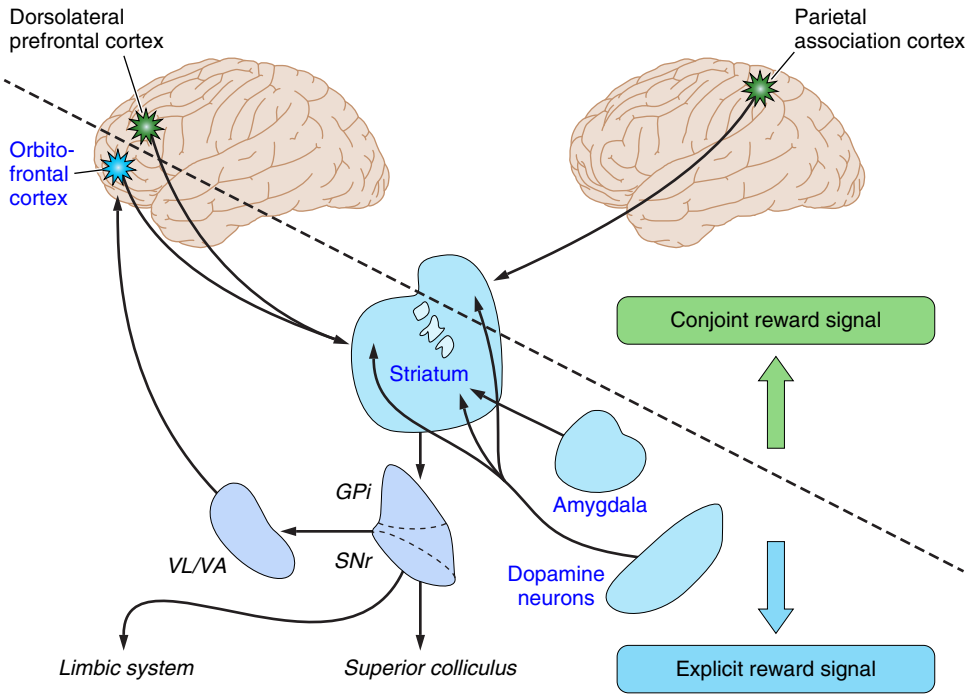


FIGURE 21. Simplified scheme of brain structures and connections involved in explicit and conjoint reward processing. The diagonal line schematically separates brain structures with neurons processing explicit reward information from structures whose neurons process reward together with sensory or action information. The striatum contains both classes.

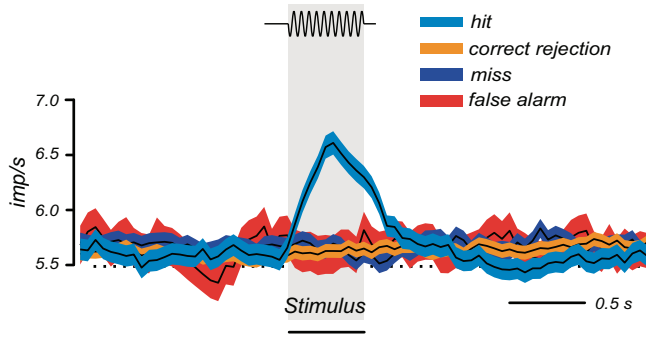


FIGURE 22. Dopamine reward response requires correct perception of reward-predicting stimulus. The reward prediction error response occurs only when the stimulus is present and is correctly reported as being present (hit; vibratory stimulus, *top*) but not when the stimulus is missed (miss) or is absent but incorrectly reported as present (false alarm). [From de Lafuente and Romo (122).]

stimulus. The responses are lower, and thus incorporate the prediction from the initial stimulus, when the animal had accurately reported stimulus presence compared with inaccurately reporting stimulus absence, even though the stimulus was physically the same. Thus dopamine responses reflect the animal's subjective perception and valuation of reward-predicting stimuli beyond their purely physical properties.

3. Distinction to unrewarding events

Reward neurons differentiate reward against neutral and aversive events. Dopamine neurons code reward but are not activated by aversive stimulus components **(FIGURE 11)** (157, 159, 160), even though their initial event detection response covaries with physical salience, novelty salience, stimulus generalization, and reward context **(FIGURE 10)**. In contrast, neurons in other reward structures respond selectively to aversive stimuli. Non-dopamine VTA neurons are activated by airpuffs and footshocks (102, 582). Some amygdala neurons are activated by airpuffs and conditioned stimuli predicting airpuffs, separately from reward neurons (422). Neurons in nucleus accumbens are differentially activated by sucrose or quinine, including their predictive stimuli (484). Neurons in orbitofrontal and anterior cingulate cortex, and TANs in striatum, respond differentially to aversive airpuffs, reward liquids, and their predictive stimuli and follow reversal between punishment and reward (10, 255, 452, 592). In dorsolateral prefrontal cortex, more neurons respond to reward than airpuff (283) and are differentially activated by gains or losses of reward tokens (532). In medial prefrontal cortex, dorsal neurons respond more frequently to air puffs, whereas ventral neurons are more sensitive to liquid reward (372). Thus punishment responses are often well separated from reward processing.

Some reward neurons respond also to punishers and thus process the motivational salience common to both reinforcers rather than reward or punishment separately. These reinforcer neurons are seen in amygdala (387), dorsolateral prefrontal cortex (283), and parietal cortex (309). In contrast to separate monotonic variations with reward and punishment magnitude, salience responses peak with high reward and high punishment and thus follow a U function (309). Thus restricted populations of reward neurons fail to discriminate against punishers and thus code motivational salience.

However, there is a caveat when interpreting the results of studies using aversive stimuli for probing negative motivational value. As indicated for dopamine responses above, it is crucial to distinguish between the physical impact of the stimulus and its motivational aversiveness. However, most studies apply presumably aversive stimuli without assessing their aversiveness to the individual animal. A good method to do so in neurophysiological studies is to estimate the negative value quantitatively on a common currency basis, using psychophysical choice procedures against rewards **(FIGURE 11*A*)** (160). Even substantial airpuffs are surprisingly not reported by the animals as having negative reward value (160).

B. Reward Value

1. Central role of subjective reward value

The ultimate function of reward is evolutionary fitness, which requires us to survive and reproduce. Thus rewards have survival value that is specific for us, and thus subjective. "Subjective" does not refer to conscious awareness but simply to the own person. Our brain detects the rewards that are necessary for this function. It is not really important what the objective, physical amount of a reward is, and often not even possible to know, as long as the reward provides us with the best chance for survival and reproduction. If we have plenty of food, a bit more does not matter much, but if we are starving, even a small amount is important. We do not necessarily go for the largest reward if it does not mean anything for us. We go for the reward that most satisfies our needs at that moment. A rat normally readily does not drink a salt solution but will do so when it is deprived of salt (472). Striving for the highest subjective value is built into our behavior by evolutionary selection. If we do not choose the highest value, we are at a disadvantage and over time fall behind and ultimately lose out against competitors that get better values. The striving for subjective value might also explain why we sometimes make choices that others do not understand. They are good for us at the moment, because we make them based on our own, private, subjective values, irrespective of whether we are consciously aware of them or not. Subjective valuation is also mechanistically reasonable. Reward function in individuals is defined by behavior driven by the brain, and the brain has evolved for our individual survival and reproduction rather than for an impersonal cause. We can identify all

substances in a wine and add up their prices from a catalog, but as a chemist and not as a wine connoisseur who is trying to maximize the personal pleasure based on subjective taste preferences. It is not the chemists's impersonal cause directed at the physical molecules but the wine connoisseur's subjective pleasure that contributes to fitness (by providing relaxation and encouraging taste discrimination training).

2. Assessing subjective value from behavior

Although subjective value is the key variable underlying approach and choices, it is a theoretical construct and not an objective measure. Utilitarianism aimed to achieve the greatest happiness for the greatest number of people (39). This might be assessed by an external person with honorable intentions (the benevolent dictator), by a hypothetical hedonimeter that measures happiness (142), or by subjective ratings (for example as "experienced utility") (259). However, benevolent dictators are implausible, hedonimeters are technically difficult, and ratings are subjective. Furthermore, there is no way of comparing subjective values between individuals because we cannot compare their feelings or, as a famous economist put it, "You cannot feel how much my tooth aches." In contrast, behavioral reactions and choices are measurable. They are often described in terms of preferences. However, preferences are internal private states and therefore prone to misjudgement and misreport. To operationalize preferences and remove their metaphysical and emotional connotations of purpose, desire, and pleasure, Samuelson (501) coined the term *revealed preferences* that are elicited by choices and follow specific axioms, notably consistency and transitivity. Thus choices elicit private and unobservable preferences hidden deep inside us and make them measurable as utility in an objective way. Other, more loosely defined, terms for value include survival value (survival of individuals or genes, referring to the reward function in evolutionary fitness), motivational or incentive value (eliciting approach behavior), and affective value (linking reward to emotion). There will be a more specific definition of economic utility further down with neuronal value signals.

There are several mechanisms for inferring subjective value from approach behavior and choices. In approach behavior, reward value is revealed by the speed (e.g., reaction time), duration, intensity, frequency, accuracy (e.g., error rate), and persistence of behavioral reactions. The reward eliciting the strongest approach behavior presumably has the highest "value." In choices, reward value is revealed by the probability of choosing one option over all other alternatives. The reward being consistently selected is considered to have the highest subjective value (assuming "rational" choices). In animals, psychophysical assessment of choices provides good measures for value. In humans, auction mechanisms allow to assess the subjective value of goods (willingness to pay). Thus approach and choices reveal the unobservable variable of value. When I see someone move faster towards an object or choose the object over another one, I get the intuitive notion that the object must have a higher value for the person. Value becomes simply a shortcut term denoting the impact of a reward on measurable behavioral choices. To avoid the circular definition when using value for explaining behavior we say that individuals behave "as if" they are maximizing value. Deducing the variable of value from behavior allows us to search for neuronal value signals for approach and choices. Such neuronal value signals would provide a physical basis for the theoretical notion of value and thus inform and validate economic choice theory.

3. Relative choices and common currency value

Behavioral choices reflect the values of choice options relative to each other. They will remain stable irrespective of value variations as long as the relative values of the options remain unchanged. For example, doubling of all option values should not affect choices. Thus value can only be expressed relative to other options within a given set of options. We can define a common reference reward, compare behavioral choices between that reference reward and any other reward, and thus establish a common value ranking or numeric scale defined by the reference reward which then serves as common currency. Money is a good example. It assesses the value of any reward in common monetary units. Any reward can be used as common currency for rewards differing in type, magnitude, probability, risk, delay, and other attributes. Just as different physical objects are measured on common scales of length or weight, different rewards can be measured on a common scale of value. Note that the common scale is only valid for an individual decision maker and does not provide a common measure to all decision makers. One dollar, or any other common currency, may constitute a different value for you than for me.

How can we choose between apples and oranges? We can compare the sensory properties of different objects using standardized scales of physical and chemical measures. As reward value is not a physical but a behavioral measure, this does not help. Rather, we should assess the values of such incommensurable rewards relative to each other. The idea of common currency allows us to quantify value relative to a unique reference that quantifies the value of rewards with vastly different properties. The activity of orbitofrontal and dopamine neurons reflects the subjective, common currency value integrated from different liquid rewards in close correspondence to behavioral choices (405) and extends to food rewards (301) **(FIGURE 23)**, suggesting a physical basis for the theoretical notion of common currency. Common currency is a necessary and sufficient feature of value coding, as it reflects independence from sensory properties and follows the idea of a reward retina as earliest neuronal detector of value (519).

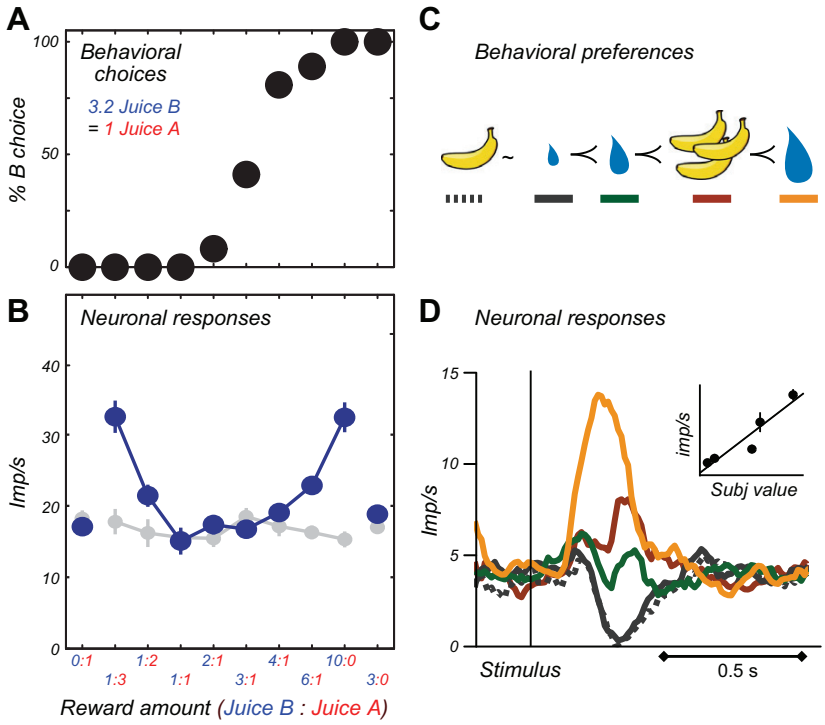


FIGURE 23. Common currency coding of subjective value derived from different rewards in monkeys. *A:* behavioral choices between two juice rewards of varying amounts reveal common scale subjective economic value. Amounts at choice indifference (50% choice of either reward) reveal 3.2 times lower value of reward B than reward A (which serves as common scale reference). *B:* common currency coding of subjective value derived from different juice rewards in single orbitofrontal neuron. Response to reward-predicting stimulus increases with amount of either juice, irrespective of juice. This activity codes the decision variable of chosen value. [*A* and *B* from Padoa-Schioppa and Assad (405). Reprinted by permission from Nature Publishing Group.] *C:* rank ordered, ordinal behavioral preferences for different liquid and food rewards, as assessed in behavioral choices between blackcurrant juice (blue drop) and mashed mixture of banana, chocolate, and hazelnut food (yellow bananas). ~, indifferent; <, preferred. *D:* common currency coding of subjective value in dopamine neurons (averages from 20 neurons). Colors refer to the rewards shown in *C*. [*C* and *D* from Lak et al. (301).]

Relative values do not always correlate with choices. Decision makers have a tendency to select the highest option unproportionally more often than its relative value against other options. Such maximizing choices favor the highest valued option irrespective of value variations as long as the ordinal value ranking among the options remains unchanged. Maximizing is reduced during exploratory behavior, when individuals occasionally check an inferior option for potential improvement, but exploration declines with increasingly stable values. These mechanisms are captured by the "temperature" parameter of the softmax function that models choices (333, 575). Exploration, modeled by higher temperature, may lead to effective decorrelation between choice probability and value and reveal serious constraints on the computation of value from choices. However, specific test methods control for these possible misreadings, including assessment of indifference points (subjective equivalence between choice options) in psychophysical procedures.

4. Value in natural rewards

Natural rewards, for which reward processing has evolved, contain nutrients and other substances of survival value. These rewards, like foods and drinks, have sensory properties, like shape, color, texture, viscosity, taste, and smell that help to detect, identify, and distinguish the reward objects. However, the sensory properties do not have nutrient value. Approach behavior and choices maximize reward value, not sensory properties. The transfer from sensory properties to value involves experiencing the value of the object, associating that value with the object's sensory properties using Pavlovian and higher forms of learning, and then using the predicted value for approach and choices. Thus sensory discrimination serves to identify the object via sensory receptors rather than determining its value. The sensory-value distinction lies at the core of reward processing; the sensory properties belong to the external object, whereas the reward properties derive from internal brain function attributing value to external objects.

Reward neurons distinguish between the sensory and value attributes of rewards. Amygdala and dopamine neurons show faster sensory and slower value responses (160, 422). Orbitofrontal and amygdala neurons follow reversed value associations of physically unchanged stimuli (422, 485). Orbitofrontal and striatal reward neurons discriminate between different food and liquid rewards but are insensitive to their visual and spatial attributes (205, 405, 601). Satiation on particular rewards reduces reward value responses in orbitofrontal and ventromedial prefrontal cortex (58, 105) but leaves sensory responses in primary gustatory cortex unaffected (487). Salt depletion induces reward responses in ventral pallidum to physically unchanged salt solutions (591). Thus the intuitive distinction between sensory and value attributes of rewards has a neurophysiological basis.

5. Non-value factors in choices

Besides being based on positive and negative economic value (gain and loss and their corresponding emotions) and exploration, real life economic choices are also determined by other factors. The factors include strategy, heuristics,

personal history, peer example, social pressure, conventions, traditions, prejudice, idiosyncrasies, morals, ethics, culture, religion, superstition, nationalism, chauvinism, daily irrationalities, and many others. Some originate from experiences that maximized gains and minimized losses and then led to automatisms that are no longer questioned for economic value. (Note that such automatisms are distinct from habits which may develop from goal-directed behavior and are based on value.) Other choice factors have evolved without explicit understanding of distant gains. Just as birds do not know the partial differential equations describing the coordination of their movements, individuals engage in behaviors without understanding why they do them, how they do them, and what the benefits are. These automatisms have evolved through evolution and provide guidance of behavior to varying but often large extents. Their advantages are transmission of beneficial behavioral traits, but their possible disadvantages are inconsistencies and domination by higher values during choices. Decisions based on automatisms would involve brain processes related to schemata and models of behavior, rules, emotions, social memory, theory of mind, and other cognitive functions.

The impact of automatisms should be considered in relation to value as key variable for economic choices. The true character of most automatisms differs fundamentally from that of gains and losses, and decision makers may not compute economic values when following habitual strategies or conventions. Then deriving value from economic choices would be very difficult and needs to be restricted to specific situations in which the automatism component can be estimated. This is often difficult and may easily lead to inconsistent choices when different situations engage different automatisms. What is truly missing is a definition of value independent of choice behavior. A neuronal value signal might provide such a measure and allow interpretations of economic choices. For example, a neuronal value signal identified independently of a particular choice situation might help to assess the contribution of value in choices in which automatisms also play a role.

C. Construction of Neuronal Value Signals

1. Value functions

Intuitively, the value of a reward increases monotonically with its physical magnitude. A larger drop of juice is worth more than a smaller drop, except when satiation results in asymptotic and then decreasing value and thus breaks monotonicity (see below). The value of a reward depends also on the frequency of its occurrence which is modeled as probability. Stimuli that predict more likely rewards are valued higher. Given their occurrence with specific magnitudes and probabilities, rewards constitute probability distributions of individual reward magnitudes, and choices between two rewards can be viewed as choices between probability distributions (417). The first statistical moment in probability distributions is the expected value (EV), which constitutes the sum of probability-weighted magnitudes (first moment of probability distributions)

$$\mathrm{EV} = \sum_i [m_i * p_i(m_i)]; \text{ over all } i \qquad (8)$$

with m as magnitude (objective property of good), p as objective probability, and i as individual reward occurrences. The mean of increasingly large samples approaches the EV.

Probability is a theoretical construct that is derived from the measured frequency of event occurrence. Probability theory, as developed by Pascal and Fermat, allowed transferring the measure of experienced event frequency into the predictive measure of probability. For biological agents and their neurons, this transfer requires learning and memory, as the frequency of events impinging on sensory receptors needs to be counted in relation to other events and expressed as probability. The fact that there are no sensory receptors for probability strengthens the notion of a theoretical construct. Thus probability assessments mediate the transition from reacting to individual events to predicting future event occurrence. The past experience of reward frequency is registered by neurons and, true to the nature of predictions, processed to predict rewards and guide future behavior. By presenting events with a given frequency to animals, we assume that this transfer takes place, and we can search for neuronal activity that reflects reward probabilities. As described above with Pavlovian conditioning, the activity is necessarily predictive of the overall frequency of reward occurrence and thus reflects the reward value of the predictive stimulus. Thus the brain reacts to and assesses the frequency of individual rewards and transfers this measure into predictive activity reflecting the probability of future rewards.

The intuition of subjective value underlying economic choices combines with the failure to explain human choices by EV (43) and led to the axiomatic formulation of Expected Utility Theory (504, 617), which defines expected utility as a measure of subjective value in analogy to *Equation 8* as

$$\mathrm{EU} = \sum_i [\mathrm{u}(m_i) * p_i(m_i)]; \text{ over all } i \qquad (9)$$

with u(m) as (subjective) utility. The main utility functions are power, logarithmic, exponential, and quadratic as well as their combinations (295), which can model human choices adequately (225). Logarithmic utility may combine the logarithmic Weber function that describes the sensory impact of reward objects with a logarithmic valuation function of objective, physical reward magnitude and probability.

Utility is a hypothetical variable reflecting the subjective preferences elicited by behavioral choices. Whereas objective, physical value is measurable directly in milliliters or grams, utility is not measurable directly but inferred from behavioral choices (and maximized by decision makers). Thus the only physical measure from which utility can be inferred is behavioral choices. Choices are the tool to assess utility, and choices are the key method that economists accept for this assessment. There is a crucial difference between subjective value and utility. Although both subjective value and utility are estimated from measured behavioral choices, utility is a mathematical function of objective value [u(m)] that allows to predict the subjective value even for choices that have not been measured. Such mathematical functions allow to determine whole distributions of subjective values and establish useful terms such as EU for comparison between choice options. Utility is a universal measure of subjective value that does not require immediate behavioral assessment every time and thus constitutes the fundamental variable of economic decision theory.

The curvature of utility functions reflects marginal utility. Marginal utility is defined as the increment in utility gained from one additional unit of consumption, or the decrement from one unit less consumption, which is mathematically the first derivative of the utility function. Progressively decreasing marginal utility leads to concave functions (downward concave, as viewed from below, decreasing first derivative), which models the decreasing welfare one derives from ever more reward **(FIGURE 24A)**. A "well-behaved" economic utility function is concave, continuous, monotonically increasing, and nonsaturating (295, 341). It is based on money of which apparently one cannot get enough. Utility functions vary between different rewards, different risk levels (631), and different individuals. Utility functions for other rewards, like foods and drinks, do saturate and may even decrease, as too much food or liquid may become aversive and thus have negative marginal utility, effectively destroying monotonicity of the utility function **(FIGURE 24B)**. The "bliss point" is the utility producing maximal satisfaction. In contrast, increasing marginal utility leads to convex functions **(FIGURE 24C)**, suggesting that wins in high ranges are considered more important than the same gains in lower ranges ("small change"). In some cases, gains are valued increasingly more but ultimately become less important, leading to an initially convex and then concave utility function (173, 339) **(FIGURE 24D)**. Effectively, all combinations of curvatures of utility functions are possible, but the typical, "well-behaved" concave utility function is the standard model in economics.

The subjective weighting of value extends to reward probability. In particular, low and high probabilities are often distorted. Prospect Theory refines *Equation 9* to

$$\mathrm{EU} = \sum_i \{\mathrm{u}(m_i) * \pi[p_i(m_i)]\}; \text{ over all } i \qquad (10)$$

with π as probability weighting function (and P replacing EU in popular notation) (258) (Prospect Theory in addition incorporates references, different curvatures, and different slopes for losses than gains). The probability weighting function $\pi(p)$ is most nonlinear at $0.0 < P < \sim 0.2$ and $\sim 0.5 < P < 1.0$ and can be modeled with a one- or two-parameter function (α) (435) with an inflection point $\pi(p) = p$ at about $P = 0.37$, or a linear-in-log-odds function (187) in which gamma reflects curvature (low = inverted S, high = regular S) and α reflects elevation. Low probabilities can be distorted by a factor of 10,000. Well-trained monkeys show similar nonlinear, inverted S-shape probability weighting as humans with similar inflections points around $P = 0.37$ **(FIGURE 25)** (559).

Given that rewards affect behavior primarily through their subjective value, and that utility u(m) constitutes the central subjective reward function in economics, we can use *Equations 9* and *10* to state reinforcement learning *Equations 1*, *2*, *4*, and *5* in terms of economic utility

$$\mathrm{UPE}(t) = \lambda(t) - \mathrm{EU}(t) \qquad (10A)$$

$$\mathrm{EU}(t+1) = \mathrm{EU}(t) + \alpha * \mathrm{UPE}(t) \qquad (10B)$$

$$\mathrm{UTDPE}(t) = \left[\lambda(t) + \gamma \sum \mathrm{EU}(t)\right] - \mathrm{EU}(t-1) \qquad (10C)$$

$$\mathrm{EU}(t+1) = \mathrm{EU}(t) + \alpha * \mathrm{UTDPE}(t) \qquad (10D)$$

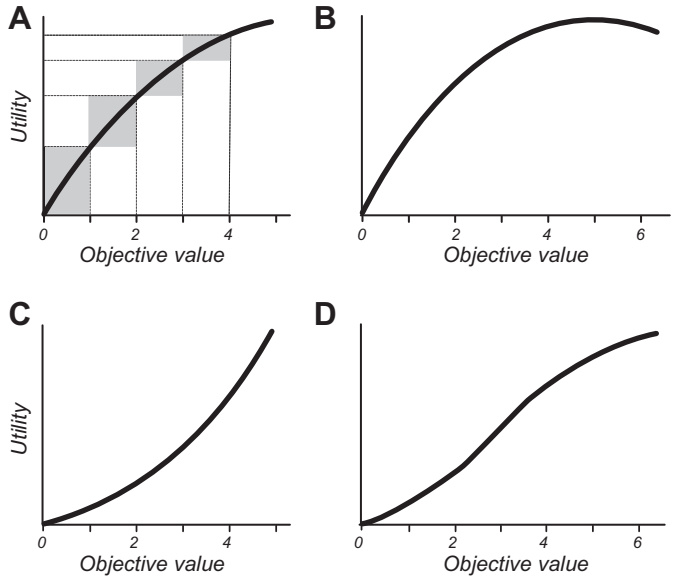


FIGURE 24. Utility functions. *A:* a "well behaved," gradually flattening ("concave"), continuous, monotonically increasing, nonsaturating utility function, typically for money. Gray rectangles indicate decreasing marginal utility (*y*-axis) with increasing wealth despite identical changes in objective value (*x*-axis). *B:* a saturating, nonmonotonic utility function typical for many biological rewards. Marginal utility becomes negative when an additional unit of consumption reduces utility (right end of function). *C:* a convex utility function, with gradually increasing marginal utility. *D:* a convex-concave utility function typical for progressively increasing marginal utility in lower ranges and progressively decreasing marginal utility in higher ranges.

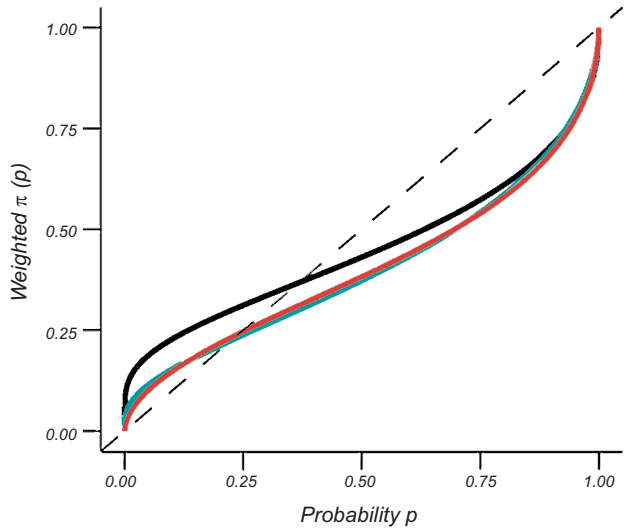


FIGURE 25. Nonlinear reward probability weighting functions estimated from behavioral choices in monkey. Fitting by one-parameter Prelec function (435) (black), two-parameter Prelec function (green), and linear-in-log-odds function (red) (187, 305). [From Stauffer et al. (565).]

Note that λ refers to the utility u(m) of the actual single reward received, whereas EU derives from the statistical utility distribution. Thus the utility prediction errors UPE(t) and UTDPE(t) constitute the key reinforcement terms for decision variables defined by economic utility.

Economists distinguish three principal forms of utility. First, scalar utility is subjective value as a mathematical function of physical, objective value. Utility can be derived from marginal utility under risk (to result in cardinal utility) or from the marginal rate of substitution without risk (amounting to ordinal utility). Cardinal utility functions established by choices under risk (617) provide a quantitative, numeric, mathematical characterization of the preference structure (257). Cardinal utility is valid up to positive affine transformations ($y = ax + b$) and can be psychophysically estimated from certainty equivalents at choice indifference between certain and risky options, starting at the distribution ends and advancing centrally by iteration (fractile or chaining procedure) (87). Cardinal utility can be expressed as u(m) (*Equation 9*) and is a requirement for establishing neuronal utility functions as neuronal signals have cardinal characteristics. In contrast, ordinal utility provides only a rank order of utilities and indicates whether a good is better or worse than another, without indicating on a numeric scale how much better or worse it is. It is only of limited value for establishing neuronal correlates of economic utility (higher versus lower neuronal responses for better or worse goods without numeric scaling). The marginal rate of substitution indicates the amount of good A a consumer is willing to give up to obtain one unit of good B while maintaining the same utility. The "distance" between curves of constant utility (indifference curves) indicates at least ordinal utility. Second, mean-variance utility is derived from objective value and risk via Taylor series expansion (see below, *Equations 21* and *21A*). Third, prospect theory extends utility by incorporating probability distortion (**FIGURE 25 AND** *Equation 10*), different loss than gain slopes (see below **FIGURE 29*E***) and reference dependency (see below **FIGURE 32** and *Equation 22A*).

The term of utility derives from the original utilitarian philosophers (39, 363). In economic decision theory, utility is defined by formal axioms (504, 617) that postulate how individuals behave in a meaningful manner "as if" they were maximizing utility. The term of utility is often used for convenience without requiring formal axiomatic tests in every instance (150, 324). This article follows this tradition, although that usage should not discourage axiomatic testing when identifying neuronal utility signals. Thus economic utility is a specific form of subjective value, which is derived from objective behavioral measures in individuals (but not across individuals). Utility has stricter requirements than subjective value by assuming an underlying, usually nonlinear, mathematical function that derives utility from objective value [u(m), as used in *Equation 9*]. If assessed from risky choices, utility is unique up to a positive affine transformation, which defines a cardinal function.

2. Neuronal value coding

The definition of cardinal, quantitative, numeric utility is important for investigating neuronal utility signals. The prime neuronal signal that is communicated to other neurons is the action potential. With sensory information, the signal arises from dedicated receptors whose stimulation induces numeric neuronal signals. Its strength as a signal is quantitatively expressed as firing rate (impulses/s) (3) and, to some extent, as pattern of action potentials. Firing rate is quasi-continuous between 0 and 1,000 impulses/s within the usual periods of 10–60 min of data sampling and usually increases or decreases monotonically with the phenomenon it codes. The same neuronal processing principles could be used for coding reward utility. For such neuronal signals to approach a meaningful mathematical function of utility, utility needs to be cardinal. It would be incorrect to derive a cardinal neuronal function from ordinal utility. This requirement would not apply if one would only look for ordinal higher firing rate for a higher (or lower) ranked, ordinal reward. With these characteristics, neuronal utility signals are physical (hardware) manifestations of a hypothetical variable that can only be inferred from behavioral choices. Although choices are the key method for assessing utility from behavior, utility neurons should also reflect utility at the time of reward reception, as this is the most important event for the survival of the organism.

Reward magnitude exerts a physical impact on sensory receptors. Reward magnitudes induce monotonically graded responses in dopamine neurons and in neurons in the stria-

tum, amygdala, and orbitofrontal, ventrolateral prefrontal, dorsolateral prefrontal, cingulate and posterior parietal cortex (12, 42, 107, 205, 262, 313, 354, 381, 405, 433, 598, 620). For rewards that do not vary in physical impact, like money, magnitude coding needs to be explained by other mechanisms including Pavlovian and cognitive processes.

Stimuli predicting specific reward probabilities induce graded responses in dopamine, striatum, and posterior parietal neurons and also in orbital, ventromedial, ventrolateral, dorsolateral, and cingulate prefrontal neurons, along with graded behavioral reactions and choices (12, 15, 17, 161, 261, 262, 354, 372, 376, 381, 402, 418, 433, 500). The transfer of dopamine responses from rewards to reward-predicting stimuli **(FIGURES 7*C* AND 14, *A* AND *D*)** may constitute a suitable correlate for the transition from reporting experienced reward frequency to coding reward probability. There are no neurophysiological studies on probability distortions. Human neuroimaging has shown probability distortions of values analogous to *Equation 10* (241, 595) but no behavioral distortions with valueless probabilities (110). Thus reward neurons transform the theoretical construct of probability into another theoretical construct called value (both of which are measurable from behavioral choices).

Increasing EVs induce graded neuronal responses in dopamine, striatum, orbitofrontal, ventromedial prefrontal, ventrolateral prefrontal, dorsolateral prefrontal, cingulate, and posterior parietal cortex neurons **(FIGURE 26)** (12, 261, 262, 354, 381, 402, 433, 418, 500, 598). In addition to characterizing reward coding, these data suggest a general biological basis for the mathematical constructs of EV, for which there are no sensory receptors. However, it is unlikely that these neurons code objective EV specifically rather than the subjective value that is monotonically related to objective value. Nevertheless, neuroscience applauds to the genius of the earlier mathematicians postulating biologically implemented theoretical terms.

Appropriate behavioral tests reveal that reward neurons typically code subjective value. Reward responses in orbitofrontal cortex, ventromedial prefrontal cortex, parietal cortex, striatum, globus pallidus, amygdala, lateral habenula and dopamine neurons closely reflect subjective behavioral choices **(FIGURE 23)** (301, 405, 433). Subjective value coding is also evident in specific situations, including satiation or deprivation without physical reward changes (58, 105, 591), differently delayed conditioned reinforcers (66, 67), temporal discounting of reward value (see below **FIGURE 28*D***) (79, 158, 285, 477, 479), and adaptation to identical reward magnitudes (see below **FIGURE 33, *A* AND *B***) (42, 284, 332, 403, 601). Details will be presented in the following sections.

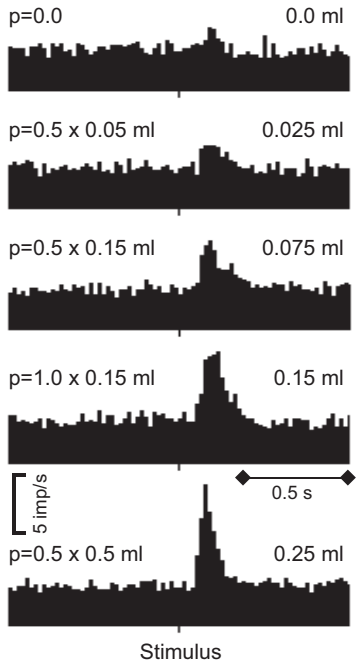


FIGURE 26. Neuronal value coding by monkey dopamine neurons. Monotonically increasing responses to increasing expected value (EV), irrespective of individual probability-magnitude combinations. EVs (*right*) are from 5 binary probability distributions with different probabilities and magnitudes of juice reward indicated at *left*. Population average from 55 dopamine neurons. Note that these responses do not suggest specific coding of EV as opposed to expected utility (EU), as at the time this distinction was not made experimentally. [From Tobler et al. (598).]

Although subjective value coding confirms the close behavioral-neuronal relationships of reward processing, it is agnostic about the relationship to physical reward value. In contrast, economic utility defines this relationship mathematically. Neuronal correlates of utility are found in dopamine neurons (560). Utility functions in monkeys may transition from convex via linear to concave with increasing reward volumes **(FIGURES 24*D* AND 27*A*)**, which reflects nonmonotonic marginal utility (first derivative of utility) and results in nonmonotonic utility prediction errors. This utility profile is closely paralleled by nonmonotonically changing dopamine prediction error responses with gambles placed at different parts of the utility function **(FIGURE 27*B*)**. With unpredicted rewards eliciting positive prediction errors, dopamine responses increase monotonically **(FIGURE 27*C*)** and thus follow the monotonic increase in utility prediction error (Δu) but not the nonmonotonic variation in marginal utility ($\Delta u/\Delta x$ or du/dx). Thus the dopamine reward prediction error response constitutes a utility prediction error signal that reflects marginal utilty but does not explicitly code marginal utility. Marginal utility relates well to prediction error, as both terms conceptualize deviations from a reference (prediction and current wealth, respectively).

Given its coding of utility, the dopamine signal can be stated as

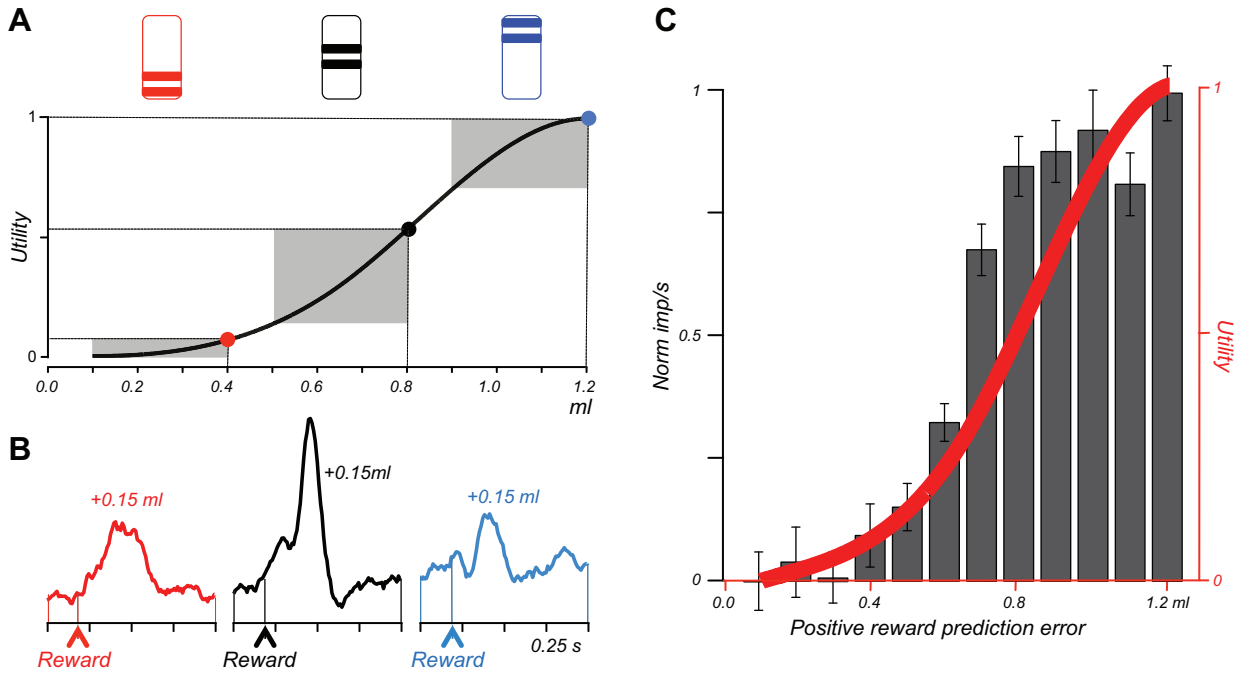


FIGURE 27. Utility prediction error signal in monkey dopamine neurons. *A*, *top*: gambles used for testing (0.1–0.4; 0.5–0.8; 0.9–1.2 ml juice; $P = 0.5$ each outcome). Height of each bar indicates juice volume. *Bottom*: behavioral utility function in monkey. Delivery of higher reward in each gamble generates identical positive physical prediction errors across gambles (0.15 ml, red, black, and blue dots). Due to different positions on the convex-concave utility function, the same physical prediction errors vary nonmonotonically in utility. Shaded areas indicate physical volumes (horizontal) and utilities (vertical) of tested gambles. *B:* positive neuronal utility prediction error responses (averaged from 52 dopamine neurons) to higher gamble outcomes in same animal (colored dots on utility function in *A*). The nonmonotonically varying dopamine responses reflect the nonmonotonically varying first derivative of the utility function (marginal utility). *C:* positive utility prediction error responses to unpredicted juice rewards. Red: utility function. Black: corresponding, nonlinear increase of population response ($n = 14$ dopamine neurons) in same animal. [*A–C* from Stauffer et al. (560).]

$$\text{DaResp}(t) \sim \text{UPE}(t) \qquad (10E)$$

and formulated for temporal difference (TD) utility error

$$\text{DaResp}(t) \sim \text{UTDPE}(t) \qquad (10F)$$

and can be fully expressed by using *Equations 6* and *7*, replacing V by EU and applying *Equations 10A* and *10E*

$$\text{DaResp}(t) \sim \lambda(t) - \text{EU}(t) \qquad (10G)$$

$$\text{DaResp}(t) \sim \left[\lambda(t) + \gamma\sum\text{EU}(t)\right] - \text{EU}(t-1) \qquad (10H)$$

As some neurons in other reward structures code also reward prediction errors, DaResp in *Equations 10E* through *10H* may be replaced by error signals from such neurons if their utility coding were established in future experiments.

As utility is not simply a measure of subjective value but derives mathematically from objective reward measures, utility coding extends the dopamine response well beyond subjective reward value coding seen with multiple rewards, effort cost, and risk (116, 156, 301, 570). Utility as higher specification of the dopamine response likely applies also to multiple rewards, effort cost, and risk whose neuronal coding is so far characterized only as subjective value (risk is already incorporated into the utility response in **FIGURE 27*B***).

3. Motivational state

Based on the intuition that the primary function of nutrient rewards derives from the need to acquire necessary substances for survival, reward value is closely related to the state of the animal. Deprivation on a particular substance enhances its subjective value, whereas satiation reduces its value. These states can be general or specific for a substance (e.g., "general" versus "sensory specific" satiety).

Satiation is captured by decreasing marginal utility in gradually flattening (concave) utility functions **(FIGURE 24*A*)**. Higher satiation leads to complete saturation and subsequent repulsion characterized by zero and then negative marginal utility **(FIGURE 24*B*)**, although such satiety does not typically occur with money. In contrast, reward deprivation is captured by the steeper part of the utility function with higher marginal utility.

Rather than directly affecting marginal utility and EU, an alternative account may conceptualize motivational states as likely nonlinear and nonmonotonic utility (EUstate) which adds to EU and results in overall EUnet, similar to other influences described below

$$\text{EUnet} = \text{EU} + \text{EUstate} \quad (11)$$

With deprivation, EUstate is positive and increases EUnet, whereas satiation turns EUstate negative and decreases EUnet. We can distinguish general from sensory specific effects on EUnet by using separate state utilities

$$\text{EUnet} = \text{EU} + \alpha * \text{EUgenstate} + \beta * \text{EUspecstate} \quad (11A)$$

with weighting coefficients α and β. General deprivation and satiety affect all rewards to some extent (EUgenstate), whereas sensory specific satiety affects primarily the particular reward (EUspecstate). The effects of deprivation and satiety on EUnet can be accommodated by the reinforcement framework by substituting EU in *Equations 10A* through *10D* by EUnet.

Satiety on individual rewards reduces behavioral reactions and neuronal responses in monkey orbitofrontal cortex to the smell, taste, and sight of rewards (105) and in ventromedial prefrontal cortex to water (58). The response decrease is specific for the satiating reward, including blackcurrant, cream, vanilline, limonène, or monosodium glutamate liquids (105). Some responses to the nonsatiating rewards remain uninfluenced, suggesting sensory specific rather than general satiety. General satiety over the daily course of the experiment reduced most responses of dopamine neurons to auditory cues for food pellets in mice (496). Opposite to satiation, salt depletion makes salt rewarding. Rats show behavioral approach to NaCl solutions after, but not before, salt depletion by furosemide while reward neurons in the ventral pallidum begin to respond to salt (472, 591). These results provide neuronal correlates for the influence of motivational states on subjective value compatible with *Equation 11*.

4. Context

Rewards occur always in an environmental context, and these contexts exert considerable influence on subjective reward value. Imagine yourself in a gourmet restaurant and hearing a loud "wow." Then imagine watching a tennis match and hearing the same "wow." Obviously the behavioral reaction to the "wow" differs depending on the context, suggesting that information from the context (restaurant versus tennis) affects the interpretation of the explicit event ("wow"). Adding a nonlinear subjective value function EUcxt captures the value gained by rewarded contexts

$$\text{EUnet} = \text{EU} + \text{EUcxt} \quad (12)$$

Animal learning theory conceptualizes the influences of context on behavioral reactions in several forms. In general, Pavlovian-instrumental transfer, instrumental (operant) learning is enhanced in the presence of a previously Pavlovian conditioned stimulus compared with an unconditioned stimulus (145). In renewal, previously extinguished conditioned responding may return depending on the contextual environment (59). In pseudoconditioning, a primary reinforcer endows the context and all its cues with reward value through conditioning. Pavlov noted "a conditioned reflex to the environment" (423) and Konorski mentioned "conditioned to situational cues" (289). Pseudoconditioning elicits approach behavior to unpaired events in rewarded contexts (367, 536, 537), as if reward properties "spill over" from the context to the explicit event. For example, in an experiment on rabbit jaw movement, the value of the unpaired stimulus was close to 0 (analogous to EU = 0), but amounted to 70% of the value of the rewarded stimulus after that stimulus had been conditioned in the same context (EUcxt = 70% of rewarded stimulus EU) (537). In all these forms, otherwise ineffective stimuli occurring in the same context gain motivational value and induce behavioral responses (191, 335), even though they have never been explicitly paired with the primary reinforcer.

Influences of contextual information are well known from everyday experiences. Labeling an odor as cheese rather than body odor, or describing flavors as rich and delicious rather than vegetable water, increases pleasantness ratings and responses in human orbitofrontal cortex (118, 189). Stating the brand name of soft drinks, like Coke and Pepsi, strongly increases preference ratings for the drink indicated and enhances prefrontal reward responses (352). In wine tasting, stated prices influence pleasantness ratings and value responses in orbitofrontal cortex (432). In risky decisions with only partly known outcomes, additional information of outcomes after the decision may reduce or enhance the value of a received outcome (329). Thus contexts add EUcxt to the EU of odors, flavors, soft drinks, and wine and gambles to modify EUnet.

5. Economic cost

Natural rewards do not only provide nutrients and calories, their acquisition usually requires calorie expenditure that reduces the gains provided by them. We need to work for them, approach them, collect them, and consume them. This intuition applies to all rewards and is conceptualized as economic cost. Every good has its cost, and we usually calculate the net benefit before handing over the money. We need to ensure that the cost does not wipe out the benefit. The value of Eve's apple is reduced by the loss of paradise.

Net benefit utility can be calculated generically as a subtraction

$$\text{EUnet} = \text{EU} - \text{EUcost} \quad (13)$$

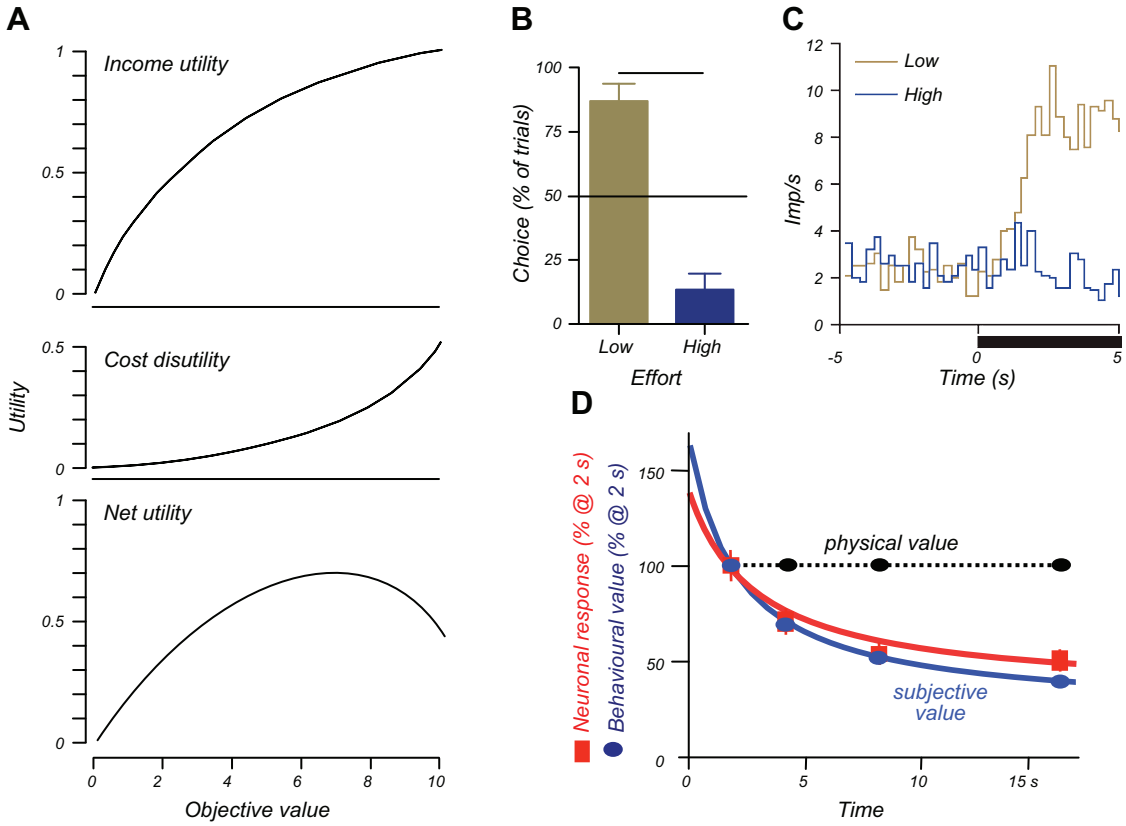


FIGURE 28. Neuronal value coding incorporates reward cost and delay. *A:* construction of net utility from reward cost. Concave income utility (*top*) minus convex cost disutility (*middle*) results in nonmonotonic net utility (*bottom*). *B:* behavioral choice preferences reveal stronger reduction of subjective reward value by high compared with low effort cost (number of lever presses) in rats (income is identical). *C:* stronger reduction of cue response of single rat nucleus accumbens neuron by high (blue) compared with low (gray) effort cost. [*B* and *C* from Day et al. (116). Copyright 2010 John Wiley and Sons.] *D:* temporal discounting in monkey choice behavior (blue) and corresponding responses of dopamine neurons (red) across reward delays of 2–16 s (hyperbolic fittings; n = 33 neurons). [From Kobayashi and Schultz (285).]

Equation 13 is based on physical net benefit, physical income, and physical cost (such as calories)

$$\text{Net benefit} = \text{income} - \text{cost} \qquad (14)$$

and becomes net utility by applying different utility functions u(income) and d(cost)

$$\text{Unet} = \text{u(income)} - \text{d(cost)} \qquad (15)$$

Income utility u(income) is often concave (gradually flattening due to decreasing marginal utility) (**FIGURE 28A**, *top*). Physical effort, as the energy expenditure required to obtaining a reward, is an economic cost, as long as it is associated with the pursuit and acquisition of that reward. As higher efforts become increasingly harder, cost is often assumed to be convex (**FIGURE 28A**, *middle*). With convex cost, the disutility of high effort may outweigh the increase of income utility, resulting in decreasing Unet and thus loss of monotonicity (**FIGURE 28A**, *bottom*). The effort cost in *Equations 13–15* needs to be distinguished from temporal discounting, which may be partly due to another form of cost, opportunity cost, and reduces reward value through a separate mechanism (see *Equations 16-18A*).

A gross simplification may neglect the different curvatures possible with *Equation 15* and apply a common logarithmic utility and disutility function

$$\text{Unet} = \log(\text{income}) - \log(\text{cost}) \qquad (15A)$$

which is equivalent to the benefit-cost ratio

$$\text{Unet} = \log(\text{income/cost}) \qquad (15B)$$

Appropriate behavioral tests would involve separate assessment of income-utility and cost-disutility functions in a common reward value currency, using choices between reward alone [to obtain u (reward)] and reward combined with cost [to obtain d(cost)], similar to composite outcomes involving punishers **(FIGURE 11A)**.

In analogy to temporal discounting (see *Equations 17* and *18* below), alternative effort models used in some neurobiological studies (419, 439) incorporate effort cost into net benefit by exponential discounting

$$\text{Net benefit} = \text{income} * e^{-k*\text{cost}} \qquad (15C)$$

to result in a utility measure

$$\text{Unet} = \text{u(income)} * e^{-k*\text{cost}} \quad (15D)$$

or by hyperbolic discounting

$$\text{Net benefit} = \text{income}/(1 + k * \text{cost}) \quad (15E)$$

$$\text{Unet} = \text{u(income)}/(1 + k * \text{cost}) \quad (15F)$$

where k is discounting factor, reflecting subjective cost sensitivity.

The pursuit of any reward reduces the chance of obtaining another reward. This opportunity cost arises with mutually exclusive rewards and is defined as the value of the highest foregone alternative reward. Thus, in principle, *Equations 13-15D* apply to every reward obtained in the presence of a viable, mutually exclusive alternative. Opportunity cost applies also when rewards are delayed and other beneficial options cannot be pursued, which is conceptualized as temporal discounting (see below).

Movements and effort may not only constitute a cost but also have intrinsic reward value (31, 358, 546, 547). A behavior may be undertaken for the pure pleasure of it and without resulting in any other reward. A kid is having fun building an airplane without getting monetary or social benefits. The intrinsic reward value can be considered as income, the required effort (and missed opportunity for homework) is the cost, and *Equation 15* becomes

$$\text{Unet} = \text{u'(intrinsic income)} - \text{d(cost)} \quad (15G)$$

with u' as utility function distinct from u. Intrinsically rewarding behavior may also lead to extrinsic reward. A mechanic loves to rebuild historic car engines and also sells them well, a typical scenario for people that love their jobs. Here, income utility derives from both intrinsic and extrinsic reward value, and *Equation 15* becomes

$$\text{Unet} = \text{u(extrinsic income)} + \text{u'(intrinsic income)} - \text{d(cost)} \quad (15H)$$

Importantly, the added value from intrinsic rewards may outweigh high effort costs. Even if u(extrinsic income) in *Equation 15H* is below d(cost), Unet remains positive, and the behavior viable, if the utility gained from intrinsic reward [u'(intrinsic income)] is sufficiently high. People that love their jobs may be content with unspectacular pay. Alternative formulations may consider intrinsically rewarding behavior as reward in its own right and apply *Equation 25* for multiple rewards (see below). Taken together, effortful behavior may both enhance and reduce net benefit. As the intrinsic gains and costs of behavior are subjective and not deducible from physical movement measures alone, it is particularly important to assess the involved utilities in choices against a reference utility using quantitative psychophysical methods. Simply assuming that movements represent only an economic cost, and estimating its utility from physical rather than behavioral measures, may be misleading.

Behavioral studies on cost follow primarily *Equation 14*. Experiments involve rats crossing barriers in T mazes (499) or working for intracranial self-stimulation (62) and monkeys performing sequential movements (60) or exchanging tokens for reward (69). Tests on humans sampling different numbers of targets comply with *Equation 15B* (108). Binary choices increase with reward value and decrease with effort cost measured as number of lever presses in monkeys (263) and rats (115, 116, 621) and lever movement against resistance in monkeys (233, 419) **(FIGURE 28B)**. These choice procedures allow comparison of subjective net benefit between the options analogous to Unet, but without explicitly assessing u(income) and d(cost). Behavioral choices are fit better by the physical cost subtraction model of *Equation 14* than by exponential or hyperbolic cost discounting of *Equations 15C* and *15E* (233). Reaction times increase with increasing effort, well fit by exponential or hyperbolic cost discounting of *Equations 15C* and *15E* (419).

Neurons in anterior cingulate cortex show activity increases with income and decreases with effort cost (233, 263), thus following the components of *Equation 14*. In contrast, orbitofrontal and dorsolateral prefrontal cortex neurons code either income or cost but not both (233). Neurophysiological responses and dopamine release evoked by reward-predicting cues in nucleus accumbens decrease with higher fixed ratio effort (115, 116) **(FIGURE 28C)**. Exponential or hyperbolic cost discounting (*Equations 15C and 15E*) affects some dopamine neurons in substantia nigra (419). Thus neuronal responses in reward structures typically reflect the influence of effort cost on reward value.

The dopamine sensitivity to effort is at odds with the failure of finding reductions in voltammetric dopamine reward responses in nucleus accumbens with effort (224, 621). However, the effort in these studies was intercorrelated with temporal discounting, which is well established to reduce reward value and corresponding electrophysiological and voltammetric dopamine responses (116, 158, 285) **(FIGURE 28D)**. Thus it is unclear why not at least temporal discounting reduced the dopamine responses. The effort-insensitive voltammetric dopamine changes were slow (lasting >8 s) and one order of magnitude lower than the typical phasic voltammetric dopamine changes [0.5–15 nM (224, 621) versus 30–70 nM (116)], which raises additional, methodological issues. The slower responses resemble dopamine changes of similar durations and magnitudes that have no correlate in dopamine impulse activity (237). Thus the slower, lower, effort insensitive dopamine changes likely do not reflect subjective value (224) but unlikley bear on the

role of phasic dopamine signals in economic value coding. Thus the slower, lower, effort insensitive dopamine changes may not reflect subjective value (224) which, however, does not bear on the phasic dopamine coding of all subjective value components tested so far, including reward amount (598), reward type (301), risk (156, 301, 560, 570), delay (116, 158, 285), and effort cost (116) all the way to formal economic utility (560).

6. Temporal discounting

Typical primary nutrient rewards, like fruits or meat, mature and decay over time. Once consumable, they risk destruction from many causes including competition by other reward seekers. Rewards often need to be consumed at particular times due of specific energy demands, and later rewards are less tangible in general. Thus specific rewards in specific situations should be harvested at specific moments. This is why we cannot wait for the pay slip, use refrigerators, drink wine at its best age, and mate only after maturation. These temporal constraints favor temporal sensitivity in the evolution of the brain's reward system and lead to personal, potentially beneficial traits of impatience, impulsivity, and risk taking. The characteristics extend to all rewards and explain the variation of subjective reward value over time. Although temporal discounting likely originates in the perishable nature of goods, it occurs even when the reward remains physically unchanged and overall reward rate is held constant, thus dissociating subjective from objective reward value.

Temporal discounting studies focus mostly on monotonic, forward decreases of value (4, 462, 473). However, value may evolve nonmonotonically in time as birds store food (443), often increasing initially and then decreasing, although the inverse is also possible. The change of reward value over time occurs between a reward-predicting stimulus and reward delivery when no operant action is required (Pavlovian conditioning) or between an action and the resulting reward. In animal learning theory, increasing temporal delays decrease the proximity (contiguity) between stimulus or action and reinforcer. Delays position the reward away from the stimulus or action and more into the background, thus reducing the subjectively perceived reward contingency (dependency of reward on stimulus or action) (175, 220). Thus both subjective devaluation and subjective contingency reduction reduce the power of reinforcers for learning.

Temporal discounting constitutes an opportunity cost when waiting prevents individuals from other activity leading to reward or from using the reward for beneficial investments. Thus temporal discounting can be stated generically as an economic cost, in analogy to *Equation 13*

$$\mathrm{EUnet} = \mathrm{EU} - \mathrm{EUtemp} \qquad (16)$$

Obtaining net benefit Unet requires two steps, assessment of u(m) from objective reward magnitudes and then its temporal discounting (296). Temporal discounting can be implemented by exponential functions, resulting in constant discount rate

$$\mathrm{Unet} = \mathrm{u}(m) * e^{-k*\mathrm{D}} \qquad (17)$$

where k is discount factor. D is subjective delay, as assessed from maximal responses at the subjectively perceived delay (peak interval procedure, Ref. 469). Alternatively, hyperbolic functions show initially higher and then gradually decreasing discounting (5, 218)

$$\mathrm{Unet} = \mathrm{u}(m)/(1 + k * \mathrm{D}) \qquad (18)$$

Hyperbolic discounting is particularly prominent against immediate rewards. When all rewards have longer delays, exponential discounting may describe the process better. The generalized hyperbola combines the two models by adding a free parameter a (325)

$$\mathrm{Unet} = \mathrm{u}(m)/(1 + k * \mathrm{D})^{a/k} \qquad (18A)$$

Unet can be assessed psychophysically with intertemporal choices between a variable early and a constant delayed reward (adjusting-amount procedure). Unet of the late reward is inferred from the utility of the early reward that produces choice indifference. Unet in monkeys decreases monotonically across delays of 2, 4, 8, and 16 s by ~25, 50, and 75%, respectively (**FIGURE 28*D***, blue) (285), with slightly better fits by hyperbolic than exponential discount functions. Temporal discounting decreases from rodents to monkeys to humans.

Some theories of temporal discounting distinguish between a prefrontal system controlling immediate, impulsive consumption and a separate valuation system (352, 538), whereas more straightforward accounts assume that a single reward value system directly discounts value according to *Equations 16-18A*, although the truth might lie in between. Some reward value coding neurons in all reward systems show decreased responses with increasing delays. Responses of monkey dopamine neurons to reward-predicting stimuli decrease monotonically across reward delays of 2–16 s (**FIGURE 28*D***, red), despite constant physical reward magnitude, thus matching closely the behavioral discounting (compare blue and red) (158, 285). Corresponding to human behavior (277), lower reward magnitudes are associated with steeper neuronal discounting (285). The same dopamine neurons code reward magnitude. Temporal discounting may reflect reduced value at the time of the predictive stimulus (input stage) or during the decision process (output stage). The reduced dopamine response occurs in imperative tasks without choices, suggesting value alterations in the input stage. The reduced neuronal response leads to reduced dopamine release in nucleus accumbens (116). Inversely, dopamine prediction error responses to reward delivery increase with longer delays. Both responses are fit better by hyperbolic than exponential func-

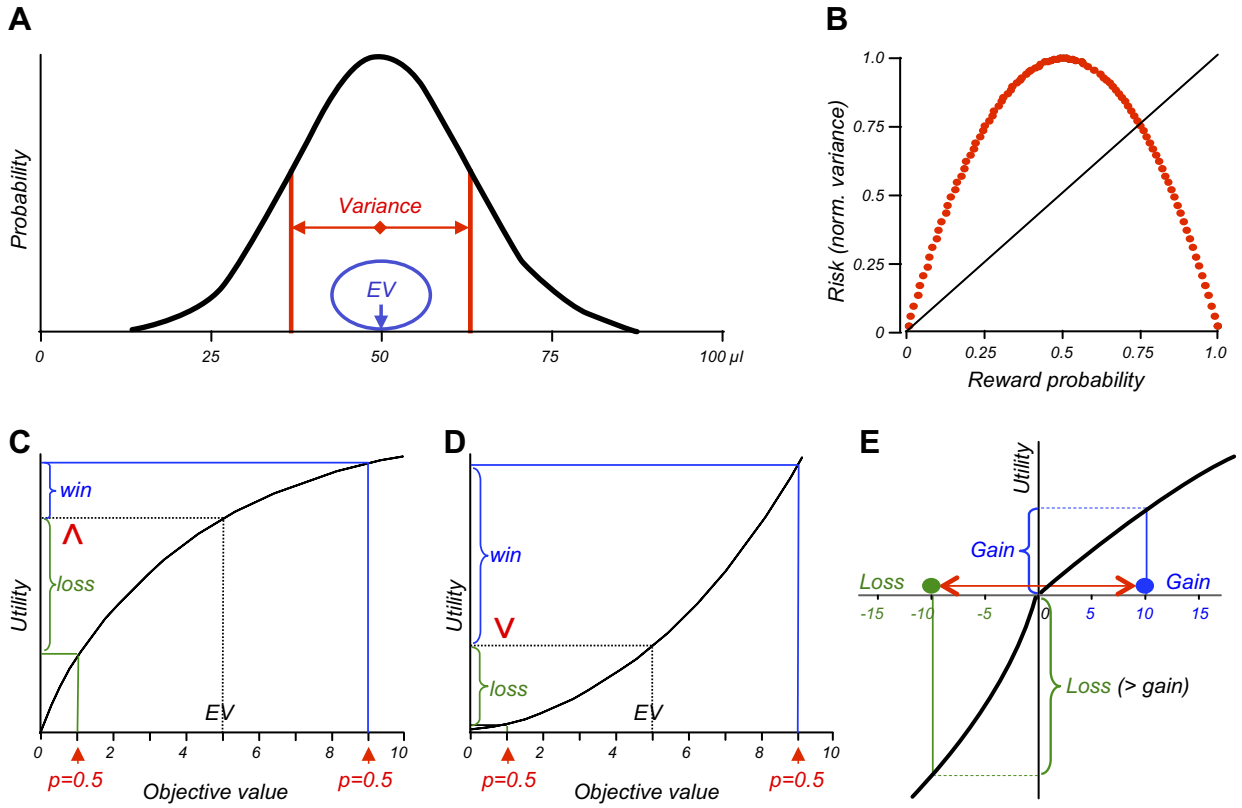


FIGURE 29. Risk and utility functions. *A:* two principal measures in probability distributions (density function). Expected value (EV, first raw moment) denotes value. Variance (second central moment) denotes spread of values and is a good measure for symmetric risk. Red arrows show ± 1 units of standard deviation (SD; square root of variance). Not included are other important risk measures, including informational entropy and higher statistical moments such as skewness and kurtosis. *B:* variance risk (normalized) as a nonmonotonic function of probability. For contrast, dotted line shows monotonic increase of value with probability. *C:* concave utility function associated with risk avoidance. The risky gamble (red, 1 vs. 9, each at $P = 0.5$) induces stronger utility loss when losing the gamble (green) than the gain when winning the gamble (blue) relative to the utility of the expected value of the certain outcome (EV = 5). *D:* convex utility function associated with risk seeking. Gain from gamble win is stronger than loss from gamble losing. *E:* explaining subjective risk notions by gain-loss functions. Due to the steeper loss slope ("loss aversion"), the loss from the low outcome of the binary gamble looms larger than the gain from the high outcome. Hence, risk is often associated with the notion of loss. [Value function from Kahneman and Tversky (258), copyright Econometric Society.]

tions. The almost perfect overlap between behavioral and neuronal discounting **(FIGURE 28*D*)** suggests that temporal delays affect dopamine responses via changes in subjective reward value. Dopamine neurons process delayed rewards as having less value.

Many reward value coding neurons in other reward structures show temporal discounting, including orbitofrontal cortex (478), prefrontal cortex (272, 475), anterior cingulate (233), premotor cortex (475), parietal cortex (477), and dorsal and ventral striatum (79, 115). Reversal of cue-delay associations leads to reversed neuronal responses, suggesting an influence of delay rather than visual properties (475). Often the same neurons code also reward magnitude (476), suggesting common subjective value coding across different rewards. Taken together, neuronal population in all major reward structures code subjective reward value during temporal discounting.

7. Risk

Rewards are, as all environmental events, inherently uncertain. The uncertainty can be formally approached via the description of rewards as probability distributions of values. In contrast to the first statistical moment of probability distributions that defines value (EV, *Equation 8*), the higher moments reflect risk, including variance, skewness, and kurtosis **(FIGURE 29*A*)**. Risk refers to the degree of uncertainty in known probability distributions, whereas ambiguity denotes uncertainty when probability distributions are incompletely known. Risk, and also ambiguity, has at least three functions in reward processing. First, individuals should perceive risk appropriately, irrespective of making a decision at a given moment. Second, as the terms of risk avoidance and risk seeking indicate, risk affects the subjective value of rewards. These two functions, and their processing by reward neurons, will be described in this section.

Third, risk informs about the spread of reward probability distributions which neurons may take into account when processing reward value within an existing distribution. This function will be described subsequently as adaptive processing.

Popular risk measures are variance

$$\text{var} = \sum_i [p_i(m_i) * (m_i - \text{EV})^2]; \text{ over all } i \quad (19)$$

where P is probability and m is magnitude. Standard deviation (SD), the square root of variance, is also a good and frequently used risk measure. EV and SD fully define Gaussian and equiprobable binary probability distributions. However, the coefficient of variation, CV = SD/EV, is often a better measure for risk in economic choices (633). Other risk measures are Shannon informational entropy (analogous to thermodynamic entropy) that seems appropriate for neuronal information processing systems (477).

Although probabilistic rewards are intrinsically risky, probability is not a monotonic measure for risk but for value (*Equation 8*). Whereas reward value increases monotonically with probability, variance as a measure for risk follows probability as an inverted U function **(FIGURE 29*B*)**. It peaks at $P = 0.5$ where the chance to gain or miss a reward is equal and maximal. Variance risk is lower at all other probabilities with more certain gains and misses.

Risk is perceived subjectively through genuine risk avoiding/seeking tendencies ("I hate/love the risk"), risk distortions ("I have no idea/am confident about the outcome"), the value at which loss/win occurs ("no gambling at high values/peanuts"), the domain, situation, and familiarity of the risky event, and hormones (86, 87, 630, 632). Some of these factors may play a role when defining subjective risk in analogy to *Equation 19* as variance of utility

$$\text{varU} = \sum_i \{\pi[p_i(m_i)] * [\text{u}(m_i) - \text{EU}]^2\}; \text{ over all } i \quad (20)$$

A fuller account would include more of the mentioned subjective factors.

It is important for the understanding of risk attitude to distinguish between objective and subjective value. The famous example is the hungry bird with challenged energy balance (87). The bird requires 100 calories to survive the night and has the choice between a certain option of 90 calories and an equiprobable gamble of 0 and 110 calories ($P = 0.5$ each). The bird would hopefully choose the gamble, as this provides it at least with a 50% chance of survival. However, from the perspective of objective value, the usually risk-avoiding bird seems to be surprisingly risk seeking, as it prefers a risky option with an EV of 55 calories over a certain 90 calories. However, the bird's behavior follows first-order stochastic dominance when we consider survival values (0 vs. 0/1 survival), which tests for meaningful and rational choices (341).

Risk avoidance and risk seeking suggest that risk affects economic choices, which can be visualized via the curvature of the utility function (*Equation 9*). With concave utility **(FIGURE 29*C*)**, the initially steeper slope indicates higher marginal utility in lower ranges. Losses move utility into the lower, steeper range and thus appear subjectively more important than gains which move utility into the upper, flatter range, resulting in risk avoidance. In contrast, with a convex utility function **(FIGURE 29*D*)**, the steeper slope in higher ranges makes gains appear subjectively more important than losses, thus encouraging risk seeking. In some cases, the curvature of utility functions can change, often from risk seeking with low values to risk avoidance with high values. Risk seeking at low values may suggest that the attraction of gambling outweighs potential losses ("peanuts effect") (152, 436, 630).

The subjective value of risk combines with riskless utility to result in the utility of risky choices, as expressed generically

$$\text{EUnet} = \text{EUrl} - \text{EUrisk} \quad (21)$$

with EUrl as riskless utility, and EUrisk as disutility due to risk. This approach recognizes the difference between riskless and risky utility functions (631). Positive EUrisk reduces EUnet and characterizes risk avoidance, whereas negative EUrisk enhances utility and characterizes risk seeking. This approach follows in kind the construction of utility from objective measures of risky outcomes (EV and variance) by the mean-variance approach in financial economics based on Taylor series expansion (238, 315)

$$\text{EU} = \text{EV} - b * \text{variance} \quad (21A)$$

with b as subjective risk weighting coefficient. *Equation 21A* applies, strictly spoken, only to quadratic utility functions.

In addition to subjective risk perceptions, the subjective effect of risk is determined by the curvature of utility functions that are typically concave **(FIGURE 29*C*)** and asymmetric for gains and losses (steeper gain functions, or steeper loss functions inducing "loss aversion") (258). Subjective pessimistic/optimistic probability distortions are further subjective influences ("today is my bad/lucky day"). These subjective factors transform symmetric variance risk into the common asymmetric notion of risk as the danger to lose (higher loss than gain utility) **(FIGURE 29*E*)**.

The most simple and controlled risky outcome is provided by a binary, equiprobable gamble ($P = 0.5$ each outcome) in which risk can be varied without changing the EV (mean-preserving spread) and which has negligible, constant probability distortion and is not skewed (497) **(FIGURE 30*A*)**. Risk attitude can be estimated by eliciting the certainty

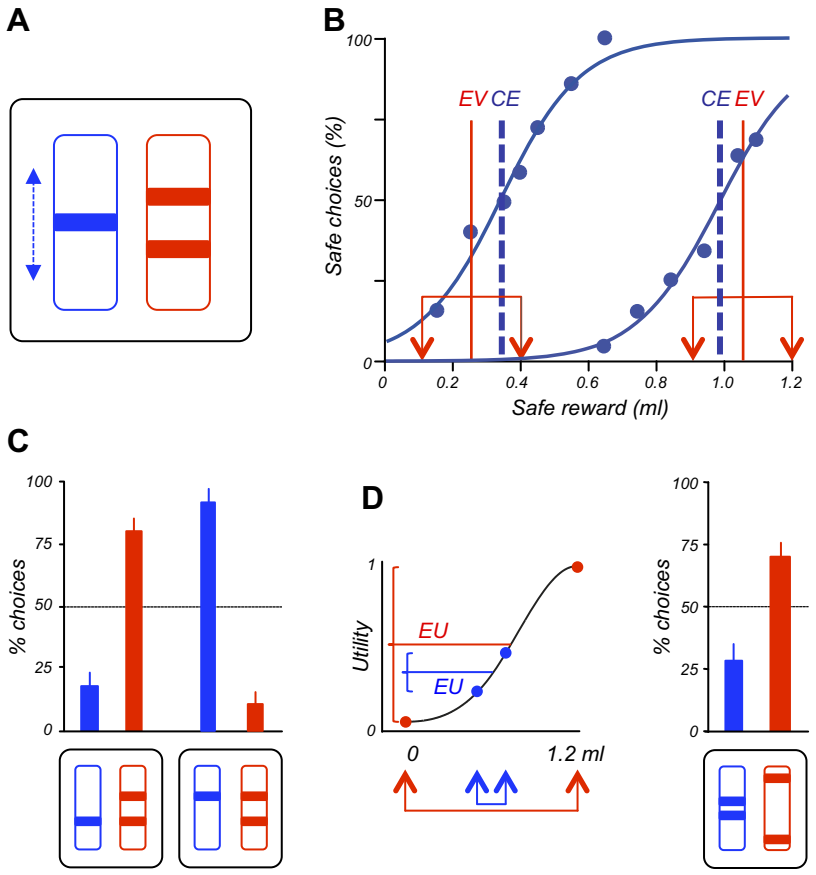


FIGURE 30. Behavioral risk measures in monkeys. *A:* stimuli indicating an adjustable certain (riskless, safe) outcome (blue) and a minimal risky binary gamble (red). Heights of horizontal bars indicate reward magnitude (higher is more). Typically, each gamble outcome occurs with probability $P = 0.5$. *B:* psychophysical assessment of subjective value of binary risky gambles by eliciting the certainty equivalent (CE). The CE is the value of the adjustable certain (riskless, safe) outcome at which choice indifference against the risky gamble is obtained (oculomotor choices in monkeys). A CE exceeding expected value (EV) indicates risk seeking (gamble at *left*), CE < EV indicates risk avoidance (*right*). Red arrows indicate the two outcomes of each gamble. *C:* choices for better options satisfy first-order stochastic dominance in monkey, indicating meaningful choice behavior. In choices against an equiprobable gamble ($P = 0.5$ each outcome), monkeys avoid the low certain reward (set at low gamble outcome, *left*) and prefer the high certain reward (set at high gamble outcome, *right*). Averaged data from binary choices involving four gambles: 0.1–0.4 ml, 0.5–0.8 ml, 0.9–1.2 ml, 0.1–1.2 ml. *D:* choices for riskier options with mean preserving spread satisfy second-order stochastic dominance for risk seeking in monkey, indicating meaningful incorporation of risk into choices. When presented with gambles in the risk-seeking domain of the utility function (left; expected utility, EU, higher for riskier gamble, red), monkeys prefer the riskier over the less risky gamble. [*B–D* from Stauffer et al. (560).]

equivalent (CE) in behavioral choices between an adjustable certain (riskless) outcome and a risky gamble. The CE is defined as the value of the certain option at choice indifference against the risky option. It is lower than EV in risk avoiders and higher in risk seekers. The risk premium reflects the value reduced or added by risk (CE < EV or CE > EV, respectively). The risk premium is usually defined as EV − CE (>0 for risk avoidance, <0 for risk seeking, 0 for risk neutrality), or sometimes as EV/CE (>1 for risk avoidance, <1 for risk seeking, 1 for risk neutrality) (558). Rhesus monkeys are often risk seeking with standard small liquid volumes (CE > EV) and risk avoiding with volumes above 0.5–0.8 ml (CE < EV) **(FIGURE 30*B*)** (355, 560, 644). Their utility function is initially convex, then linear and ultimately concave **(FIGURES 24*D* AND 27*A*)**, thus showing an inflection as conceptualized and found in human utility functions (152, 339, 436, 630).

Formal tests of stochastic dominance (341) confirm that monkeys understand the risky stimuli (560). First-order stochastic dominance helps to define meaningful choices without requiring assessment of a utility function. However, higher EV does not necessarily satisfy first-order stochastic dominance, due to nonlinear utility, nonlinear probability weighting, and risk attitude. In statewise dominance, a case of first-order stochastic dominance, a gamble dominates another gamble if all its outcomes are equal except at least one outcome that is better. Monkeys' behavior satisfies this criterion in choices between a certain reward and an equiprobable, binary gamble whose low or high outcome equals the certain reward. They avoid the low certain reward (**FIGURE 30*C***, *left*) and prefer the high certain reward to the gamble (*right*). Second-order stochastic dominance tests whether risk is incorporated into subjective reward value in a consistent fashion. Monkeys' choices satisfy also this test. With mean preserving spreads, they prefer more risky over less risky gambles in the lower, risk-seeking range of reward volumes **(FIGURE 30*D*)** and show stronger risk preference against certain outcomes (391). Also, monkeys prefer risky over ambiguous options (207). Thus monkeys make meaningful choices under risk consistent with expected utility theory.

The orbitofrontal cortex plays a key role in risky choices (317). Human patients and animals with lesions in orbitofrontal cortex or nucleus accumbens show altered risk and ambiguity sensitivity distinct from value (35, 88, 174, 240, 268, 364, 370, 445). Correspondingly, specific orbitofrontal neurons in monkeys show responses to stimuli predicting binary, equiprobable gambles **(FIGURE 31*A*)** (391). These risk responses occur irrespective of visual stimulus properties and only very rarely reflect reward value, which is coded in separate orbitofrontal neurons. The absence of value coding argues also against the coding of salience, which is associated with both risk and value. Some monkey orbitofrontal neurons are activated by unsigned (absolute)

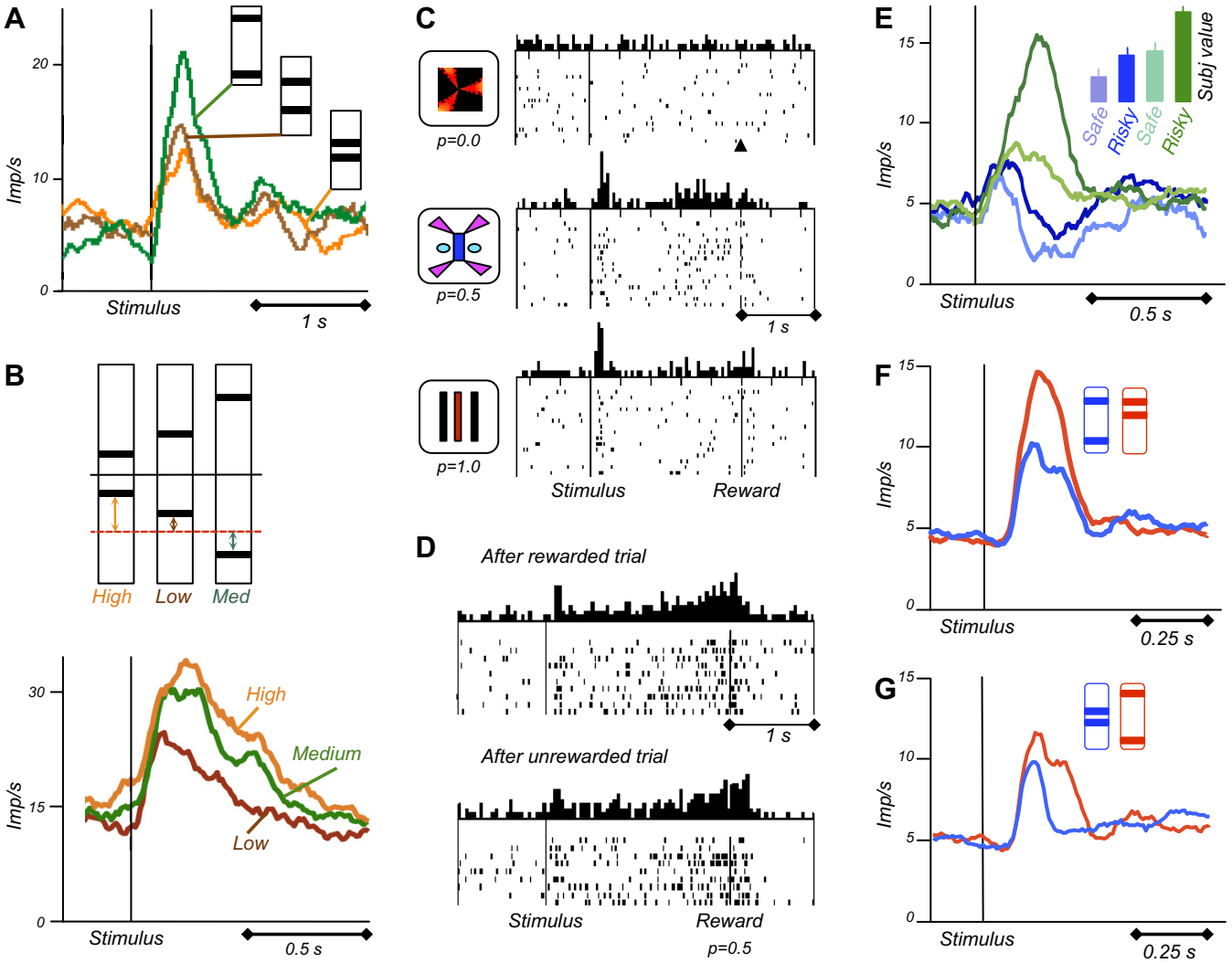


FIGURE 31. Neuronal risk processing in monkeys. *A:* coding of risk in single neuron of orbitofrontal cortex. Height of bars indicates liquid reward volume, two bars within a rectangle indicate an equiprobable gamble (P = 0.5 each). The three gambles have same mean but different variance (9, 36, and 144 ml × 10^{-4}) (mean-preserving spread). This neuron, as most other orbitofrontal risk neurons, failed to code reward value (not shown). [From O'Neill and Schultz (391), with permission from Elsevier.] *B:* coding of risk prediction error in orbitofrontal cortex. *Top*: risk prediction error (colored double arrows), defined as difference between current risk (vertical distance between bars in each gamble indicates standard deviation) and predicted risk (common dotted red line, mean of standard deviations of the three gambles). Colored double vertical arrows indicate unsigned (absolute) prediction errors in variance risk. *Bottom*: averaged population responses from 15 orbitofrontal neurons to unsigned risk error. Such signals may serve for updating risk information. [From O'Neill and Schultz (392).] *C:* risk coding in single dopamine neuron during the late stimulus-reward interval. The activation is maximal with reward probability of P = 0.5, thus following the inverted U function of variance risk shown in **FIGURE 29*B***. Trials are off line sorted according to reward probability. [From Fiorillo et al. (161).] *D:* independence of risk activation in single dopamine neuron from reward outcome in preceding trial. Activation is not stronger after positive prediction errors (*top*) than after negative prediction errors (*bottom*). Thus prediction error responses do not backpropagate in small steps from reward to stimulus. Both rewarded and unrewarded current trials are shown in each graph. [From Fiorillo et al. (162).] *E:* influence of risk on subjective value coding in dopamine neurons. The population response is stronger with subjectively higher valued certain (safe) juice rewards (green: blackcurrant; blue: orange; both 0.45 ml), demonstrating value coding. This value response is enhanced with risky juice volumes (equiprobable gamble of 0.3 and 0.6 ml, same mean of 0.45 ml), corresponding to the animal's risk seeking attitude (higher subjective value for risky compared with certain rewards, *inset*). [From Lak et al. (301).] *F:* dopamine responses satisfy first-order stochastic dominance, demonstrating meaningful processing of stimuli and reward value under risk (averages from 52 neurons). In the two gambles (blue, red), the higher outcomes are equal, but the lower red outcome is higher than the lower blue outcome, defining gamble dominance. *G:* dopamine responses satisfy second-order stochastic dominance, demonstrating meaningful incorporation of risk into neuronal utility signal (averages from 52 neurons). The red gamble has greater mean-preserving spread than the blue gamble, defining dominance with risk seeking. [*F* and *G* from Stauffer et al. (560).]

risk prediction errors, defined as unsigned difference between current and predicted risk, which may serve for updating risk information (FIGURE 31B) (392). In a categorical discrimination task, responses in orbitofrontal neurons reflect the riskiness of olfactory stimuli but in addition differentiate between correct and erroneous choices in correlation with decision confidence (264). The orbitofrontal risk signal is distinct from movement-related activity (153), some of which correlates with particular, asymmetric forms of salience (393). In human prefrontal cortex, BOLD signals show direct risk coding, distinct from value (599), which reflects risk perception irrespective of choice. Outside of orbitofrontal cortex, supplementary eye field neurons differentiate categorically between certain and risky trials and code certain and risky reward value differentially (551), and septum neurons show slow risk responses closely resembling the risk ramp of dopamine neurons (373).

Dopamine neurons code reward risk during the stimulus-reward interval after the value prediction error response (FIGURE 31C) (161). The activation ramps up to the reward and varies with probability in an inverted U function, analogous to risk measures such as variance, standard deviation, and entropy. It increases with variance and standard deviation, even when entropy is kept constant. Thus the signal reflects risk rather than value (FIGURE 29B). An analogous signal is also seen in the human ventral striatum, which is distinct from value and most likely reflects synaptic activity induced by the electrophysiologically measured dopamine risk signal (437). Although its latency is too long for immediate decision processes, the dopamine risk signal might affect the subsequent prediction error signal at the time of the reward. In coding risk, the signal varies with surprise salience and thus may affect the learning constant of attentional (associability) learning theories (425, 336). Also, the slower risk signal might contribute to the risk sensitivity of the faster, phasic dopamine signal (see below) (156, 301, 560, 570).

Straightforward implementations of TD models may produce similar ramps by backpropagation of prediction error responses from reward to the preceding conditioned stimulus in small steps via imaginary stimuli (374, 524), thus resembling the dopamine risk response (388). As the backpropagating activation would be small and dispersed, it would appear as ramp only when averaging across multiple trials (183, 388). However, the dopamine risk signal occurs in single trials without averaging and does not backpropagate via imaginary stimuli, as suggested by its absence after prediction errors (FIGURE 31D) (161, 162), but jump in single steps from reward back to the preceding actual stimulus (FIGURE 14A) (117, 410, 598). The ramp is not a necessary feature of TD models but occurs only in specific implementations (374, 388, 524). Biologically more plausible TD implementations and spectral timing models use stimulus eligibility traces, as originally suggested (FIGURE 14B) (574), and model well the single step backward transition of dopamine prediction error responses from reward to the next preceding conditioned stimulus without error backpropagation and thus without a ramp during the stimulus-reward interval (FIGURE 14, C AND D) (71, 410, 573). In particular, models using a biologically consistent striatal and midbrain circuit replicate well the dopamine risk ramp (581). Thus the dopamine ramp represents a genuine risk response and not the necessary byproduct of TD models.

In addition to the late and slow dopamine risk signal, risk affects the phasic dopamine value response. The subjective reward value signal of dopamine neurons is enhanced when monkeys prefer risky over certain outcomes of low or intermediate reward magnitudes (FIGURE 31E) (156, 301), suggesting incorporation of risk into the final construction of utility in possible compliance with *Equation 21A*. Correspondingly, dopamine concentration changes following risky cues are reduced in risk avoiding rats and enhanced in risk seekers (570). In formal tests informed by economic theory (341, 497), dopamine responses follow first- and second-order stochastic dominance (FIGURE 31, F AND G), suggesting meaningful coding of economic value and risk and incorporation of risk into neuronal substrates of subjective value. Thus three types of experiment demonstrate that dopamine neurons incorporate risk appropriately into neuronal substrates of subjective value and formal utility, namely, value responses modified by risk in correspondence with risk attitude (FIGURE 31E), second-order stochastic dominance between different risky outcomes (FIGURE 31G), and processing of utility assessed via CEs across risk seeking and risk avoiding domains (FIGURE 27B) (560).

Risk influences also subjective value coding in orbitofrontal cortex (444, 482) and modifies saccadic activity in posterior cingulate cortex (355). In human prefrontal cortex, risk affects value responses in close correspondence to individual risk attitudes in analogy to *Equation 21A* (596). Direct risk coding for risk perception and incorporation of risk into subjective value constitute two different processes that allow multiple uses of risk information for behavior.

Taken together, the distinct neuronal coding of value and risk suggest that Pavlovian reward predictions concern probability distributions defined by value and risk. In contrast to reward magnitudes that impact directly on sensory receptors, the coding of EV, probability, and risk of reward is not explained by peripheral sensory stimulation but requires central neuronal mechanisms including memory. Irrespective of these implementation details, the separation of reward signals according to the statistical moments of probability distributions, and the possible correspondence to Taylor series expansion, provides a biological basis for the mathematical treatment of rewards as probability distributions, demonstrates the biological plausibility and imple-

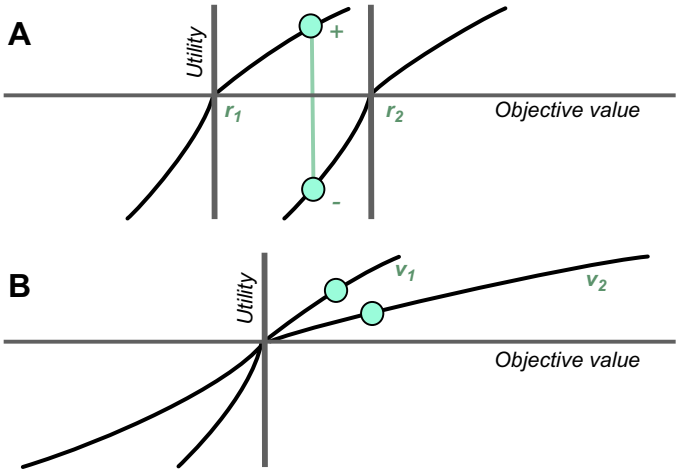


FIGURE 32. Scheme of adaptive reward processing. *A:* adaptation to expected value (r_1, r_2): shift of reference point changes utility gain to loss for identical objective values. *B:* adaptation to variance (v_1, v_2): slope adaptation to variance results in larger utility for lower objective value. [Value function from Kahneman and Tversky (258), copyright Econometric Society.]

mentation of these mathematical concepts, and incorporates them into the neuronal basis of learning theory.

8. Adaptive processing

Whereas the number of neurons and their maximal firing rate limits the processing capacity of the brain, the number of possible rewards is huge. Dedicating equal processing capacity to all possible rewards would lead to low slopes of reward-response functions and thus poor reward discrimination. However, reward availability varies considerably between situations. An efficient discrimination mechanism would restrict neuronal processing to the rewards that are currently available in a given environment, similar to sensory adaptation to probability distributions of ambient visual intensities (148, 235, 307).

For optimal response slopes and thus best use of the limited neuronal coding range, reward processing may adapt to all statistical moments of reward distributions, including EV and variance **(FIGURE 29*A*)**. When schematically separating these two moments, changes in EV lead to processing shifts with maintained response slopes **(FIGURE 32*A*)**, whereas variance affects response slopes, which become steeper with narrower distributions and flatter with wider distributions **(FIGURE 32*B*)**. These adaptations lead to optimal reward discrimination with steepest possible response slopes while maintaining as wide reward coverage as possible.

The consequences of adaptive processing are changes in net utility, which may be expressed generically by adding a term (EUadapt) that summarily covers adaption to all statistical moments of the current probability distribution of objective values

$$\mathrm{EUnet} = \mathrm{EU} - \mathrm{EUadapt} \qquad (22)$$

This approach includes reference-dependent valuation in prospect theory derived from *Equation 10*

$$\mathrm{EU} = \sum_i \{u(m_i - r) * \pi[p_i(m_i)]\};\ \text{over all } i \qquad (22A)$$

with r as reference value (258). The term r is replaced by EVr if reference r derives from a probability distribution rather than being a single value

$$\mathrm{EU} = \sum_i \{u(m_i - \mathrm{EV}r) * \pi[p_i(m_i)]\};\ \text{over all } i \qquad (22B)$$

Equation 22B describes adaptation to the first moment of the probability distribution of objective values. The adaptation to EV can lead to vastly different utilities for identical objective values **(FIGURE 32*A*)**.

A milder approach to reference-dependent valuation combines nonadapted and adapted utility (290). This approach would refine the generic *Equation 22* to

$$\mathrm{EUnet} = \alpha\mathrm{EU} + \beta(\mathrm{EU} - \mathrm{EUadapt});\ \alpha + \beta = 1 \qquad (22C)$$

α and β are additional weighting coefficients. In the terminology of Köszegi and Rabin (290), EU represents consumption utility, and (EU-EUadapt) is reference dependent gain-loss utility. Thus EUnet consists of weighted absolute and relative utilities. Considering nonadapted value recognizes the fact that we know cognitively the absolute reward value, even though we value it relative to references or discount it over time. We know that \$100 is still the same \$100 irrespective of having received \$50 or \$500 the last time around, even though we may feel more elated receiving them as a positive than negative surprise.

Adaptive processing may not be limited to a single reference value and should be extended to full distributions. Inclusion of other statistical moments such as variance adds slope changes. We extend *Equations 10* and *22B* by introducing ω for slope adaptation to variance (or standard deviation)

$$\mathrm{EU} = \sum_i \{u[\varpi * (m_i - \mathrm{EV}r)] * \pi[p_i(m_i)]\};\ \text{over all } i \qquad (22D)$$

A convenient term for ω is 1/standard deviation (384, 438). The adaptation to variance can lead to larger utilities for identical or even lower objective values **(FIGURE 32*B*)**.

Adaptations to reward probability distributions underlie several interesting behavioral phenomena. The negative contrast effect derives from a downshift of reward and reduces behavioral reactions and preferences compared with the same low reward occurring without preceding higher values (51, 104, 593), even when the lower outcome adds value (513). A positive but less prominent contrast effect exists also. The opponent process theory of motivation ac-

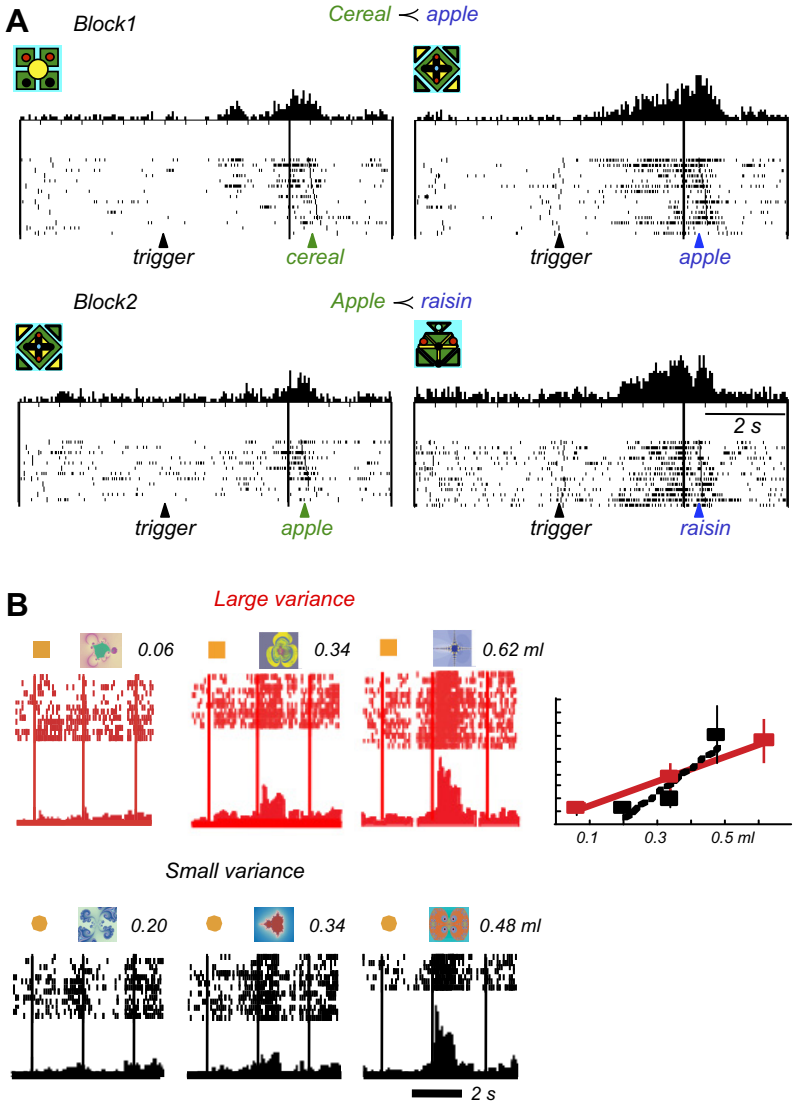


FIGURE 33. Adaptive neuronal reward processing in monkey orbitofrontal cortex. *A:* response adaptation to approximate expected subjective value in single neuron. Trial block 1 offers pieces of cereal or apple, separate block 2 offers same apple (indicated by same stimulus) or raisin. Behavioral choices ranked cereal < apple < raisin. Visual instructions predict type of reward, trigger stimuli elicit arm movement leading to reward. [From Tremblay and Schultz (601).] *B:* response adaptation to variance of reward amount (ml) in single neuron. *Inset* shows change of reward-response regression slope in same neuron. [From Kobayashi et al. (284).]

counts for observations indicating that cessation of reward is aversive, whereas end of punishment is rewarding (552). This effect may well derive from short-term adaptation to the initial reward or punishment, respectively. Adaptations to distributions may also explain the dichotomy in humans between "experienced utility" reflecting utilitarian, hedonic value (39) and "decision utility" measured in overt choices, although overweighting of recent and maximal value is also important (259). This effect is also seen in monkeys (52). The distinction may explain irrational, suboptimal, dominated choices. Monkeys favor [one food morsel plus a 50% chance of a second morsel] over [two food morsels with 50% chance of losing one], although EVs are both 1.5 morsels, although endowment effect and loss aversion may also play a role (91).

Reward responses in dopamine neurons, striatum, amygdala, parietal cortex and orbital, dorsolateral, and cingulate prefrontal cortex adapt to reward magnitude and reward probability distributions as suggested by *Equation 22* (137, 628). More formally, the adaptations may occur to changes in EV (**FIGURE 33*A***) (42, 106, 232, 601), variance (**FIGURE 33*B***) (284, 420, 598), or both combined (80, 403). In subtracting reward value from the mean (prediction error) and dividing it by standard deviation, dopamine neurons code a z-score of reward value (598). Divisive normalization from sensory physiology describes well the adaptations to changes in individual values of choice options in reward neurons of parietal cortex (332). The ventral striatum is importantly involved, as lesions here reduce reward contrast in rats (326). Taken together, the adaptations would be driven by neuronal EV and risk signals and result in a match between the probability distribution of neuronal responses and the probability distribution of current reward values. The result would be dynamically varying, optimal reward response slopes and thus efficient reward processing for choices. The loss of the actual values would be immaterial in choices comparing the options relative to each other. Thus, in adapting to the two key moments of reward probability distributions, reward neurons make ef-

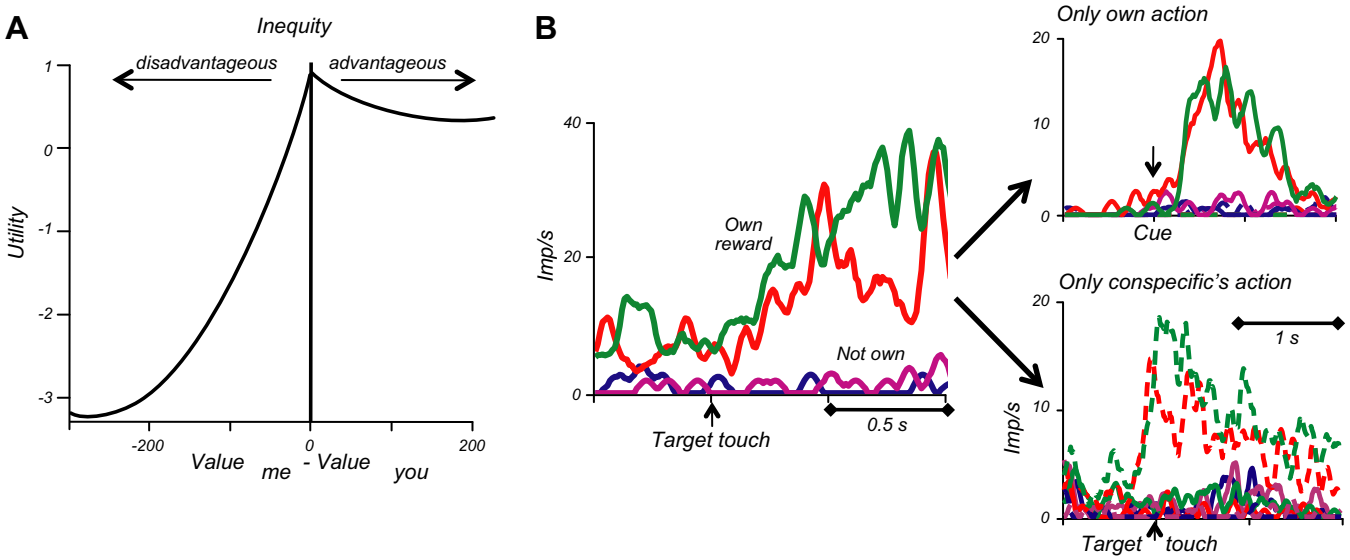


FIGURE 34. Social reward processing. *A:* social reward inequity aversion. Difference between own and other's reward reduces my utility. Usually, disadvantageous inequity (getting less than others, negative difference, *left*) has stronger and more consistent effects than advantageous inequity (getting more than others, *right*). [From Loewenstein et al. (324).] *B:* action-dependent coding of social reward in single striatal neurons. In the imperative reward giving task (modified dictator game), two monkeys sit opposite each other across a touch-sensitive computer monitor and give reward to itself and the conspecific. *Left*: neuronal activation when receiving own reward, either only to itself (red) or together with conspecific (green), but no activation with reward only to conspecific (violet) or to nobody (blue). *Right*: activation with own reward occurs only with own action (*top*, continuous lines) or only with conspecific's action (*bottom*, dotted lines, different neurons between *top* and *bottom*). [From Báez-Mendoza et al. (25).]

ficient use of processing resources for optimal reward discrimination and choice.

The EU shift between environments and the slope adaptation may result in lower utility of higher absolute reward values **(FIGURE 32*B*)**. Thus adaptation may lead to violation of rank order, menu invariance, independence from irrelevant alternatives, and transitivity, which are crucial for meaningful economic choices. However, at least two processes may serve to maintain transitivity in the presence of adaptive valuation. First, the mentioned predictive EU and risk signals in dopamine neurons (598) and orbitofrontal cortex (391) convey information about the two key moments of the reward probability distribution of each environment. These signals may combine with the adapted value signal to generate an accurate, absolute, nonadapted value signal and thus maintain transitivity at the neuronal level. Second, adaptations require time and may only concern subpopulations of reward neurons. Indeed, only a fraction of orbitofrontal reward neurons show adaptation, whereas the remainder codes reward magnitude irrespective of other rewards (403, 406, 284). Although longer test periods might have increased the number of adapting neurons, some orbitofrontal reward neurons seem to code nonadapted, absolute value necessary for maintaining transitivity. Adaptation and absolute value coding by different neuronal populations would combine efficacy with transitivity.

9. Social rewards

Observation of social partners and comparison of rewards between them is crucial for competition and cooperation that improve performance and give individuals access to otherwise unobtainable resources and ultimately enhance fitness. Social factors and contexts have at least two distinct reward functions.

First, viewing and encountering others is a reward on its own. Rhesus monkeys sacrifice quantities of liquid reward for viewing body parts of conspecifics, demonstrating that social images have reward value (124). Neurons in monkey parietal cortex code the reward value of conspecific's faces and hindquarters (279), although orbitofrontal and striatal neurons code these social stimuli often separately from liquid rewards (629).

Second, reward to others affects own reward processing. With benevolent exceptions ("Mother Teresa" and many other mothers), individuals hate to receive less reward than others when all other factors are equal. Disadvantageous inequity aversion typically induces a sharp, gradually declining, drop in the utility of own reward **(FIGURE 34*A*)** (324). Emotional correlates are envy and distaste for unequal or unfair outcomes. Individuals often also dislike, usually to a lesser extent, receiving more reward than others. This advantageous inequity aversion may elicit a similar

but weaker utility drop. Associated emotions are compassion, sense of fairness, guilt, dislike of inequity, positive reciprocity, group welfare, "warm glow," and maintaining personal social image (83, 149). The aversion may turn into advantageous inequity seeking (getting more than others) with competition and other advantage seeking situations associated with pride, sense of achievement, superiority feelings, and retaliation. Good behavioral tools for assessing inequity are the ultimatum game in which rejection of unfair offers assesses disadvantageous inequity aversion, and the dictator game in which the reward fraction handed to the opponent reflects advantageous inequity aversion (83, 149).

Final own expected utility in social settings EUnet can be stated generically as

$$\mathrm{EUnet} = \mathrm{EU} - \mathrm{EUsocial} \quad (23)$$

EU stands not only for income utility (*Equation 10*) but also for any utility EUnet derived from reward components (*Equations 11, 12, 13, 18, 21*, and *22*). EUsocial is derived from the difference between own value x and value x_o received by the opponent(s), weighted by disadvantageous (α) and advantageous (β) inequity coefficients in analogy to a popular model (150), resulting in

$$\mathrm{EUnet} = \mathrm{EU} - \alpha * \max\left[(x_0 - x), 0\right] - \beta * \max\left[(x - x_0), 0\right] \quad (23A)$$

where α and β are disadvantageous and advantageous inequity coefficients, respectively. Alternative accounts focus on reciprocal "fairness weights" which enhance own reward value when being nice to a nice opponent and reduce own value when being mean to a mean opponent, thus stabilizing cooperation and disfavoring unsocial behavior (442). Both models suggest that reward to others has appreciable value for the observer. Although condensing social reward value into a common value variable is tempting and would help to conceptualize decisions, nonvalue factors such as strategy may importantly influence decisions (50, 150, 151).

Chimpanzees and rhesus, cebus, and marmoset monkeys show prosocial tendencies and actively observe rewards in conspecifics (77, 123, 359). Cebus show disadvantageous but no advantageous inequity aversion (69). Rhesus prefer receiving own reward alone over both animals receiving the same reward, suggesting advantageous inequity seeking (negative β of *Equation 23*), and they prefer seeing conspecifics receive reward over nobody receiving reward, suggesting disadvantageous inequity seeking (negative α) in this setting (23, 89). Disadvantageous inequity reduces the effort cebus are willing to spend (612). Many of these social effects may depend on the relative social positions of the monkeys within their group and would merit further investigation. These social tendencies seem simplistic compared with those of humans but may constitute basic building blocks for social encounters and exchanges.

Reward neurons in monkey orbitofrontal cortex and striatum process primarily own rewards in social settings, whereas neurons in anterior cingulate cortex distinguish between own and other's reward or sense primarily conspecific's reward (**FIGURE 34*B***, *left*) (23, 25, 89). Striatal reward neurons distinguish between the social agents whose action leads to own reward (**FIGURE 34*B***, *right*) (25). These neurons are only active when either own action or the other's action results in own reward, and many of them do not show this activity with a computer instead of a monkey opponent. Neurons in medial frontal cortex show activations during observation of behavioral errors of conspecifics, sometimes together with observation of reward omission (647). Some premotor mirror neurons distinguish between reward and no reward irrespective of who is receiving it (78). These neuronal signals mediate the distinction between own and others' rewards and actions and thus convey basic components of social reward function. When testing social competition, reward neurons in prefrontal cortex differentiate between competitive and noncompetitive video game performance (234). In humans, disadvantageous inequity reduces, and advantageous inequity activates, ventral striatal reward responses (165), reflecting α and β of *Equation 23A* (the coding direction may reflect different signs of the two coefficients or different neuronal coding slopes). Taken together, social reward neurons code fundamental components of competition and cooperation.

10. Multiple reward attributes

Single rewards come seldom with a single attribute. Attributes include income utility, cost, delay, risk, and social interaction quantified by *Equations 9-23A*. Their aggregate subjective reward value can be condensed into a single variable called total utility. This has advantages. First, extracting the variable once from the multiple reward attributes and then using it for choices would save the decision maker computing time in crucial moments and thus enhance evolutionary fitness, even if occasional attribute changes require partial recalculations. Second, the brain could transform the condensed neuronal value responses into signals for specific decision variables for competitive decision mechanisms such as winner-take-all (see below).

Total utility EUtotal of a single reward with multiple attributes can be derived from a combination of attributes, each with their own expected utility derived from likely nonlinear and possibly nonmonotonic utility functions

$$\mathrm{EUtotal} = \mathrm{EU} + f_i * \mathrm{EUstate} + f_2 * \mathrm{EUcxt} - f_3 * \mathrm{EUcost} - f_4 * \mathrm{EUdiscount} - f_5 * \mathrm{EUrisk} - f_6 * \mathrm{EUref} - f_7 * \mathrm{EUsocial} + \mathrm{INTER} \quad (24)$$

where f_1 to f_7 are weighting coefficients reflecting the different contributions of each reward attribute. In some cases, the attributes can be so closely linked to individual rewards that the total utility u(total) of a reward can be directly derived from adding or subtracting their utility u(att) to the income utility u(m)

$$\text{u(total)} = \text{u}(m) - \text{u(att)} \quad (24A)$$

Establishing a full probability distribution from u(total) of all rewards would allow computation of EU with *Equations 9* or *10* that amounts to the same EUtotal as stated in *Equation 24*.

Take the example of a hamburger in a pub and apply *Equations 10, 11, 12, 13, 18, 21, 22*, and *23*. Its expected utility EU is defined by the probability distribution of the utilities of the slightly varying hamburger sizes u(burger). My hunger enhances EU by EUstate, and the familiar pub context associated with food and drink adds to the burger's attraction with EUcxt. However, the burger does not come for free (EUcost) and takes a moment to cook because of a busy kitchen, subjecting it to temporal discounting (EUdiscount). Burger quality varies, to which I am averse, thus reducing its EU by EUrisk, although it is still the best food in the pub, which puts its EU above a mean reference and adds EUref (in a gourmet restaurant it would have lower rank and thus lose EUref). When the burger finally arrives, it is smaller than my neighbor's and thus elicits disadvantageous inequity aversion, but better than my dieting other neighbor's salad, inducing mild advantageous inequity aversion, which both reduce the burger's EU by EUsocial. Some of these attributes are known to interact which is captured summarily by INTER. For example, smaller rewards are discounted steeper (lower EUdiscount with smaller EU) (277) and larger stakes increase risk aversion (lower EUrisk with larger EU) (152, 630).

11. Multiple rewards

Most rewards are composed of several individual reward objects whose values add up and may interact. Our hamburger meal is composed of the burger with its calories and taste, a tomato salad with taste and appealing visual aspects, and a pint of ale with water, calories, taste, and alcohol, thus totaling eight individual rewards. However, ale is sometimes poorly pulled (not in my pub), which would constitute an aversive component that would need to be subtracted from the positive components to obtain a summed final reward. If utilities of several rewards are assessed separately on a common scale, they can be weighted and added up to a summed utility, with possible interaction (INTER)

$$\text{EUsum} = \sum_i (f_i * \text{EUtotal}_i) + \text{INTER}; \quad i = 1,\ n \text{ rewards} \quad (25)$$

with f_i as weighting coefficients. Each of the rewards is likely to have its own saturation dynamics. INTER models multisensory integration as interaction that can make aversive taste or odor in some, usually small quantities add to reward value, rather than subtract from it, like pepper on meat, spicy mustard on sausage, or quinine in sugary fizzy drinks.

Equation 25 can be used for assessing the value of an otherwise unquantifiable reward on a scale defined by a common reference reward. For example, in choices between [juice plus picture] versus [juice alone], the value of the picture is reflected by the sacrificed amount of its associated juice to obtain choice indifference (124). Punishers can be included into *Equation 25* where they reduce EUsum. They constitute negative value but not economic cost, as they do not reflect energetic, temporal, or monetary expenditures. Behavioral choices of monkeys reveal the value summed from rewards and punishers compatible with *Equation 25* **(FIGURE 11*A*)** (160).

With the specification of final utility EUsum, the reinforcement *Equations 10A–D* can be restated as

$$\text{UPEsum}(t) = \lambda(t) - \text{EUsum}(t) \quad (26)$$

$$\text{EUsum}(t+1) = \text{EUsum}(t) + \alpha * \text{UPEsum}(t) \quad (26A)$$

$$\text{UTDPEsum}(t) = \left[\lambda(t) + \gamma \sum \text{EUsum}(t)\right] - \text{EU}(t-1) \quad (26B)$$

$$\text{EUsum}(t+1) = \text{EUsum}(t) + \alpha * \text{UTDPEsum}(t) \quad (26C)$$

Analogous definitions can be made for utility prediction errors of multiple attributes of single rewards by replacing EUsum by EUtotal of *Equation 24*. With multiple rewards and reward attributes, summed, single prediction errors of *Equations 26* and *26B* would not require separate computations of utility prediction error for every reward and attribute. Such prediction errors can be used by economic decision mechanisms for straightforward EU updating by standard reinforcement rules.

Alternatively, prediction errors may be computed from separate, possibly multi-attribute, rewards that are added up in a weighted manner, in analogy to *Equation 25*

$$\text{UPEsum}(t) = \sum_i \left[f_i * \text{UPEtotal}_i(t)\right] + \text{INTER}; \ i = 1,\ n \text{ rewards} \quad (26D)$$

$$\text{UTDPEsum}(t) = \sum_i \left[f_i * \text{UTDPEtotal}_i(t)\right] + \text{INTER}; \ i = 1,\ n \text{ rewards} \quad (26E)$$

A similar calculation is used for incorporating fictive outcomes into gamble values (329). This computation takes the prediction error of each reward into account. An analogous breakdown of prediction error computation, and hence updating, can be made for individual reward components listed in *Equation 24*.

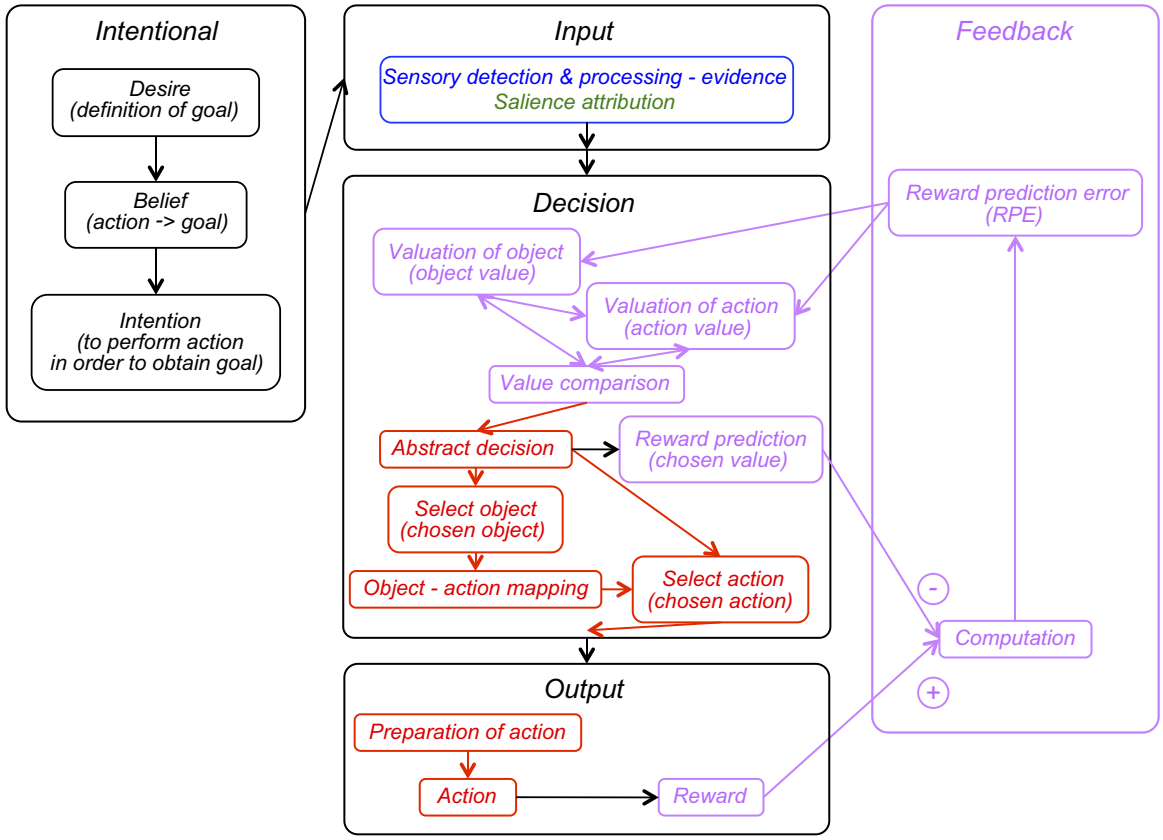


FIGURE 35. Steps and components of voluntary economic decisions. Five component processes occur in sequence, and to some extent in parallel depending on the characteristics of the decision. Decision variables are indicated in central box below their respective mediating processes (object value, action value, chosen value, chosen object, chosen action).

Given that the dopamine prediction error signal codes utility (560), the dopamine responses (*Equations 10E* and *10H*) can be restated for multiple rewards by replacing EU by EUsum

$$\mathrm{DaResp}(t) \sim \lambda(t) - \mathrm{EUsum}(t) \qquad (26F)$$

and for the temporal difference error

$$\mathrm{DaResp}(t) \sim \left[\lambda(t) + \gamma \sum \mathrm{EUsum}(t)\right] - \mathrm{EUsum}(t-1) \qquad (26G)$$

DaResp in *Equations 26F* and *26G* would apply also to utility error signals in other brain structures.

Common scale value coding is a prerequisite for applying *Equation 25* to neuronal signals and is seen in orbitofrontal cortex with different liquid rewards (405) and images of body parts (629), in parietal cortex with body part images (279), and in dopamine neurons with different liquid and food rewards (301) **(FIGURE 23, *B* AND *D*)**. Neuronal signals reflect the sum of positive and negative values from liquid or food rewards and aversive liquids or air puff punisher in monkey dopamine neurons **(FIGURE 11*D*)** (157, 160) and anterior cingulate cortex (10), and from odor rewards in human orbitofrontal cortex (190). Compatible with coding negative value as component for *Equation 25*, some frontal cortical neurons are depressed by losses of gained rewards (532). Ideally, neuronal investigations would test EUsum with all constituent components, although this is impractical and experimenters usually investigate only a few of them at a time.

D. Economic Decision Mechanisms

1. Component processes of decisions

In a highly simplified scheme, economic decisions may involve five component processes **(FIGURE 35)**. The initial intentional component comprises the desire of obtaining a reward (*left*). This emotion is based on the knowledge, or at least a hunch, that there is a reward that is worthwhile to pursue. With a belief that an action can be attributed to the reward, an intention can be formed to perform an action to get the reward. Thus the reward becomes the goal of the intentional action. In the practicalities of laboratory experiments, these processes can be driven by explicit, temporally specific stimuli indicating the availability of rewards or arise spontaneously by processes internal to the decision maker based on environmental information. The desire to get a

beer may be driven externally by the sight of a pub or internally by the heat of a summer day.

Once the intentional stage has been completed, the inputs to the decision mechanism are processed (**FIGURE 35,** *top*). Perceptual decisions require stimulus detection and then evidence accumulation for identifying the sensory reward component, tracking the object and assessing its associated salience components (**FIGURE 1,** *left* and *middle*). Correct identification of each reward option is the basis for choosing the best one. For example, perceptual decisions identify motion direction from the degree of dot motion coherence; the percentage of correct identification translates into reward probability (134, 389).

The perceptual stage is followed by valuation of the individual reward objects (**FIGURE 1,** *right*) or, depending on the characteristics of the decision, by valuation of the actions required to obtain each reward object. Valuation comprises the immediate identification or gradual accumulation of value evidence, similar to the accumulation of sensory evidence in perceptual decisions. The process may either parallel the gradual identification of the object or action or start only after they have been completely identified. The decision stage (**FIGURE 35,** *middle*) begins with object and action valuation and comprises the selection of the best reward by comparing instantaneously or gradually the values of the available objects or actions. Valuation and value comparison may evolve partly in parallel (double pink arrows). Purely sensory and motor aspects no longer play a role, unless they affect value, such as effort cost (*Equation 13*). Following valuation and value comparison, the abstract decision specifies the result of the process. It concerns only the decision itself and leads to three selections: the winning object, the winning action, and the prediction of the reward value extracted from the winning object values and action values. If the decision primarily specifies the winning object, subsequent object-action mapping determines the action required to obtain the reward object.

Obtaining the selected object and performing the selected action requires motor decisions for preparing and executing the necessary action (**FIGURE 35,** *bottom*), as every decision will ultimately lead to an action for acquiring the chosen reward. The quality of motor decisions determines the efficacy of the economic decision process. Decision makers want to get the action right to get the best reward. In cases in which different possible actions have the same cost and lead to the same reward, motor decisions may become independent of economic decisions. In these cases, the choice of the particular action may depend on noise or on motor exploration. Apart from equivaluable actions, it is unclear how much motor decisions can be independent of economic consequences. If actions are chosen for their own sake, they have intrinsic reward value and thus are economic decisions. After experiencing the selected reward object, its value is compared with the value predicted during the decision process (**FIGURE 35,** *right*). The resulting prediction error is conveyed to the valuation process in the input stage where it serves to update object values and action values.

The scheme of **FIGURE 35** should not suggest strictly sequential processing of decision variables that transitions from sensory identification via value assessment to decision-making. Real life situations are often characterized by noisy evidence, unstable values, moving objects, and changing actions and thus are likely to engage several of these processes in parallel. Furthermore, dopamine, striatum, amygdala, frontal, and parietal neurons show not only hugely different but also considerably overlapping reward valuation and decision activity. Several arguments suggest nonsequential processing of evidence, value, and decisions. First, dorsolateral frontal and posterior parietal neurons are involved in spatial movement processes but also show considerable reward value coding **(FIGURES 20*E* AND 21)** (313, 381, 433, 627). Second, ramping activity for evidence accumulation is the hallmark of cortical perceptual decisions but occurs also in subcortical structures during reward expectation and valuation (18, 40, 215, 221, 260, 523). Third, abstract decision coding and chosen action coding overlaps with reward valuation in subcortical structures, including amygdala (188) and striatum (79, 275, 306). Although the different processes may involve different neuronal pools within the same brain structures, the data nevertheless question a strictly sequential processing scheme and suggest more interactive and overlapping processing of evidence, value, and decision-making within the different structures and through subcortical loops.

In addition to many nonvalue factors that determine economic choices (see above), even purely value-based decisions may not involve the complete or correct consideration, computation, or comparison of values. A first important deviation from proper value processing is due to noise. Noise can be derived from uncertain sensory evidence (experimentally modeled, for example, by partly coherent random dot motion), varying salience attribution, imprecise estimation of economic value (economic noise, for example, when estimating probabilities from changing frequencies, or when deriving value from these probabilities), and noisy learning constants inducing unstable value predictions. Except for sensory evidence, which is external to the brain, the noise results from neuronal noise in sensory receptors and subsequently engaged neurons. Ways to deal with noisy information are averaging out the noise and accumulating the evidence over time. A second deviation from proper value processing occurs with exploration. Options that were previously suboptimal might have improved in the meantime and should better be recognized to prevent missing the best options and ensure utility maximization in the long run. Exploration is captured by ε-greedy and softmax functions (575). A single parameter models the proportion

of exploration and dominated choices (higher ε and softmax "temperature" for more exploration). Thus utility maximization involves a healthy mix of exploitation (choosing known best options) and exploration (choice of dominated but potentially better options). The proportion of exploration should be fine tuned to the dynamics of the environment to result in maximal reward. It should be higher in volatile (higher ε or temperature) compared with more stable situations (lower ε or temperature). A third deviation from proper value processing occurs with satisfying when individuals fail to consider all available options and choose only from a subset of many options that seem to satisfy their immediate desires (85, 545). Individuals may stop searching for higher value and thus make suboptimal choices. However, in some cases, the cost of considering all options may be high and reduce substantially the net benefit (*Equations 13–15*), in which cases satisfying may constitute the optimal behavior. A fourth deviation from proper value processing may be due to adaptive value processing that includes inaccessible and thus irrelevant options. In these cases, irrelevant alternatives do affect option values and may make less valuable options appear more valuable and direct choices towards them. The common result of the various forms of incomplete value consideration may be selection of suboptimal, dominated options whose values are lower than those of the best option, a behavior that compromises utility maximization.

2. Decision variables

A decision variable is what the decision maker tries to control and use to achieve the goal. It is the crucial independent variable driving the decision process. The basic decision variable in economic choices is subjective value and, more specifically, formal economic utility. The basic assumption in all economic decision theories is that decision makers aim to maximize utility. Utility is derived from several attributes including reward magnitude, probability, motivational state, context, cost, delay, risk, reference, and social interaction, as stated in *Equations 9–25* above. This notion of decision variable differs from the more technical definition that includes all sources of inputs to decision rules and mechanisms (185).

An additional definition of decision variables derives from the specific decision mechanism that serves to control them. Decision variables should be appropriate for this mechanism. Economic decision variables constitute the inputs to the decision mechanism and reflect the value of each option irrespective of the actual choice (**FIGURE 35,** *top*). The value is expressed as formal economic utility. It concerns specifically the value of a reward object (object value) and the value obtained by an action (action value). Although the decision process cannot change what is on offer, it maximizes the value by selecting the best object or action from the existing inputs (**FIGURE 35,** *middle*). This process requires sensory evidence for identification of each option as prerequisite for subsequent valuation of the options. Sensory evidence itself is not a decision variable in economic decisions, as it is imposed by the environment, and controlling it would not necessarily lead to the best reward. Furthermore, decision variables are closely related to the outputs of the decision process. The outputs controlled by the decision maker concern the reward value predicted from the decision (chosen value), the object that conveys that value (chosen object), and the selection of the valuable action (chosen action). Preparation, initiation, and execution of the action for obtaining that value are downstream processes (**FIGURE 35,** *middle* and *bottom*).

The value in the decision variables is rarely innate but acquired and updated using various forms of learning, in particular Pavlovian processes, operant habit learning, and operant goal-directed behavior, as described above. Through these three forms, objects and actions acquire formal economic utility, using model-free reinforcement learning as most basic mechanism, as stated in *Equations 10B, 10E, 26A*, and *26C*. More complex learning and updating of utility may use various forms of model-based reinforcement learning.

Perceptual decision models, which are crucial for the sensory identification of reward objects (**FIGURES 1,** *left*, and **35**, *top*), assume the gradual accumulation of evidence that needs to be extracted from noisy stimuli. Two main types of perceptual decision models vary in the way the inputs are compared with each other. Random walk, drift, or diffusion models have a single integrator with separate decision thresholds specific for each option **(FIGURE 36*A*)**. Evidences are continuously compared between options through mutual inhibition at the input level (254, 302, 447, 454, 549, 565). The resulting difference signal grows and diverges over time towards one of the decision thresholds. The decision is executed for the option whose threshold is reached first. In contrast, race models have distinct accumulators for each option, each one with its own decision threshold. The evidences grow separately until the first option reaches a specific threshold, which then generates the corresponding behavioral choice **(FIGURE 36*B*)** (327, 614, 622). Race models are appropriate for choices involving several options that are tracked separately until a final comparison can be made. Intermediate forms between diffusion and race models differ in architecture and parameter setting (55). The valuation of each option (**FIGURES 1,** ***right*****, and 35,** ***top***) may occur as soon as the perceptual decisions have been made and each reward object has been identified. In other cases, the value may not be immediately apparent and requires some time to evolve into a stable representation. Only then can an economic decision variable arise. Thus a general formulation of economic decision mechanisms should include accumulation processes for sensory and value evidence.

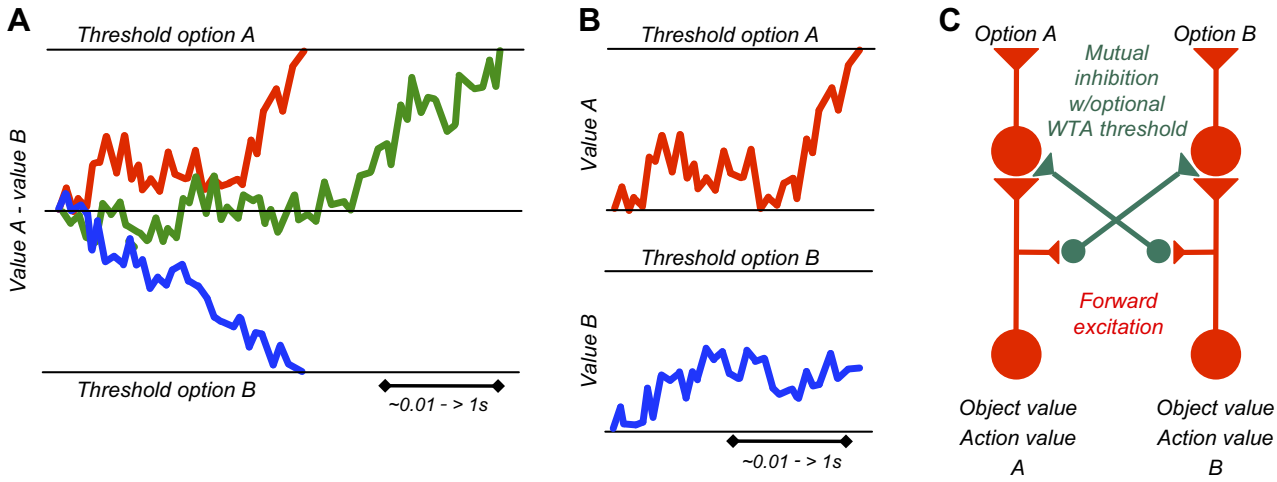


FIGURE 36. Schematics of decision mechanisms. *A:* diffusion model of perceptual decisions. Difference signals for sensory evidence or value increase in a single integrator towards a specific threshold for one or the opposite choice option (red, blue). Later threshold acquisition results in later decision (green). Time basis varies according to the nature of the decision and the difficulty in acquiring sensory evidence and valuing the options. *B:* race model of perceptual decisions. Separate signals for each choice option increase in separate integrators towards specific thresholds. The option whose value signal reaches the threshold first (red) will be pursued, whereas the lower signal at this point loses (blue). *C:* basic competition mechanism. Separate inputs for each choice option (*A*, *B*) compete with each other through lateral inhibition. Feedforward excitation (red) enhances stronger options, and mutual inhibition (green) reduces weaker options even more, thus enhancing the contrast between options. This competition mechanism defines object value and action value as input decision variables. For a winner-take-all (WTA) version, a threshold cuts off the weaker of the resulting signals and makes the stronger option the only survivor.

The basic mechanism of the decision process is competition between independent options and underlies a large variety of decision models (55, 56, 441, 585, 623). It would mediate the on-going comparison and divergence of evidence in the diffusion models **(FIGURE 36*A*)**, achieve the final comparison between options in the race models **(FIGURE 36*B*)**, or constitute a decision mechanism on its own when evidence is immediate and accumulation negligible. The minimal model comprises two inputs that map onto two outputs while inhibiting each other's influence **(FIGURE 36*C*)**. Through forward excitation coupled with mutual lateral inhibition, the stronger neuronal signal becomes even more dominant by not only inhibiting more the weaker signal but also by being less inhibited by the weaker input. The mechanism amplifies the graded differences between the inputs, analogous to lateral inhibition of sensory systems, and, with various alterations, forms the basis for competition in a wide variety of diffusion and race models for perceptual and motor decisions (55, 56, 441, 585, 622, 623) and may also apply to economic decisions. Recurrent excitation would generate gradually increasing evidence accumulation in components of the network (622, 623). An additional threshold produces a winner-take-all (WTA) mechanism by removing the weaker signal and turning the graded difference into an all-or-none output signal that reflects only the value of the strongest option or consists of an all-or-none decision signal for the winning option. This comparative mechanism provides definitions for the major decision variables.

Neuronal decision signals should reflect the decision variables defined by the decision mechanisms. The art is to identify formal, biologically plausible decision models that employ the mechanisms defining the decision variables and are implemented with identifiable neuronal activities during various trial epochs **(FIGURE 20, *A* AND *D*)**. The activities concern the initial accumulation of evidence, implemented by gradually increasing ramping activity in perceptual decisions (**FIGURES 35, *top*, and 36, *A* AND *B***), and the subsequent valuation of the options. The central competitive process and its outputs (**FIGURES 35,** *middle* and *bottom*, and 36*C*) are reflected in distinct neuronal activities coding the input decision variables of object value (405) and action value (500), the abstract decision output (188), and the chosen value (405), chosen object (38), and chosen action (543). Updating of decision variables via prediction errors occurs globally via dopamine neurons and locally through specific groups of non-dopamine reward neurons (**FIGURE 35,** *right*). Neuronal decision signals are typically investigated in choice tasks, although basic processing aspects may be studied in imperative, forced-choice trials, which allow to assign neuronal coding to individual choice options and to predictions from a single stimulus. The following sections are devoted to these descriptions.

3. Evidence and valuation

The initial stage of the decision process concerns the rapid or gradual acquisition of evidence and the valuation of the

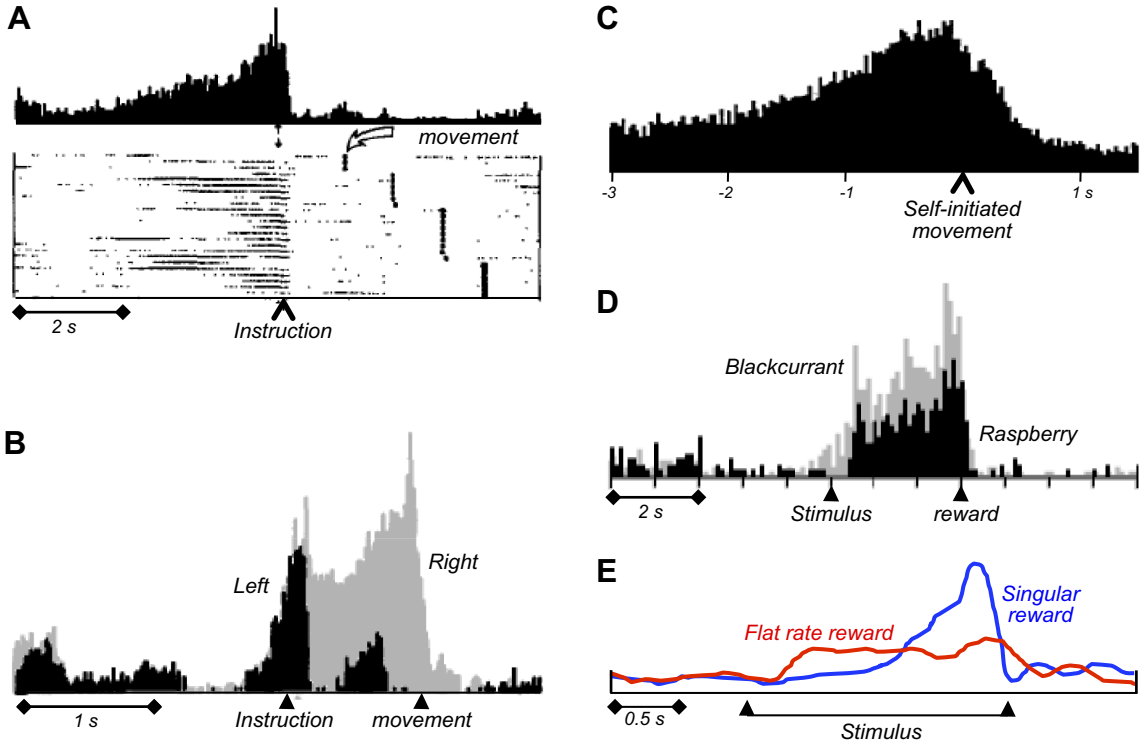


FIGURE 37. Neuronal ramping activity preceding stimuli, action, and reward. *A:* gradual activity increase during the expectation of an initial visual instruction stimulus in premotor cortex neuron. Open arrow points to movement triggering stimulus. [From Mauritz and Wise (350), with kind permission from Springer Science and Business Media.] *B:* gradual increase of left-right differential activity during movement preparation in premotor cortex neuron. Activity terminates just after the movement triggering stimulus and before the movement. Left (black) and right (gray) refer to target positions. [From Kurata and Wise (298).] *C:* ramping activity preceding self-initiated movements in striatum neuron. Activity begins gradually without external imperative stimuli and terminates with movement onset. [From Schultz and Romo (528), with kind permission from Springer Science and Business Media.] *D:* differential reward expectation activity in striatum neuron. Activity is lower in anticipation of raspberry juice (black) than blackcurrant juice (gray). [From Hassani et al. (205).] *E:* reward expectation activity in amygdala neuron reflecting instantaneous reward probability. Activity ramps up steeply before singular reward occurring at predictable time after stimulus onset (increasing reward expectation), but shows lesser increase before reward occurring pseudorandomly during stimulus (flat reward expectation with flat occurrence rate). [From Bermudez et al. (40).]

identified options (**FIGURE 35**, *top*). Acquisition of evidence takes time, in particular when the evidence itself is noisy. Neurons in many brain structures have the capability to bridge time gaps and accumulate information through persistent activity that may ramp up to future events. Ramping activity is one of the major neuronal mechanisms underlying decisions. This form of neuronal activity concerns sensory information in perceptual decisions, motor processes in action decisions, and economic utility with all its components in economic decisions (*Equations 9–13, 16*, and *21–25*).

Ramping activity reflects a general propensity for persistent, gradually increasing activity in many cortical and subcortical neurons. It occurs without choice during several processes. During the expectation of sensory events, activity increases gradually toward predictable visual stimuli in premotor cortex (**FIGURE 37*A***) (350), prefrontal cortex (65), parietal cortex (248), and striatum (6, 18, 215, 526, 528). During the instructed preparation of eye and arm movements, activity ramps up and differentiates between movements in primary motor cortex (467), premotor cortex (**FIGURE 37*B***) (298, 491, 493, 636), frontal eye field (74, 506) [and its projection neurons to superior colliculus (147)], supplementary eye field (101, 506), supplementary motor area (491, 493), parietal cortex (109, 184), superior colliculus (379), and striatum (**FIGURE 20*F***) (526, 528). Ramping activity to a threshold occurs also several seconds in advance of internally chosen, self-initiated movements in neurons of premotor cortex, supplementary motor area, parietal cortex, and striatum; these neurons are typically not activated with externally triggered movements (**FIGURE 37*C***) (299, 311, 337, 380, 397, 491, 493, 526, 528). Similar human encephalographic ramping activity resembling the readiness potential is the basis for an accumulator decision model for movement initiation (515). During reward expectation, ramping activity occurs in orbitofrontal cortex (216, 544, 602), dorsal and ventral striatum (**FIGURES 32*C* AND 37*D***) (18, 205, 215, 523), and amygdala (**FIGURE 37*E***) (40). The anticipatory ramping activity seen with sensory events,

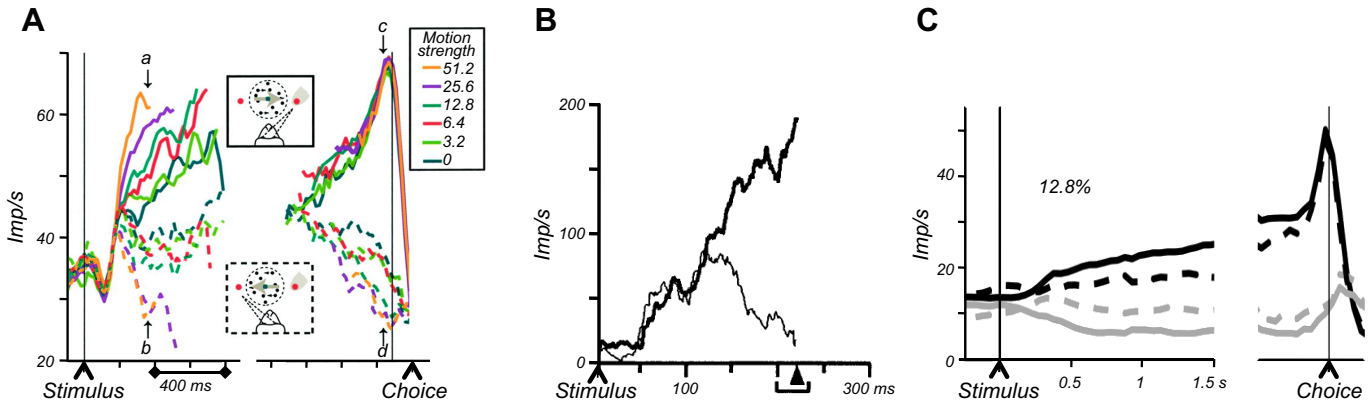


FIGURE 38. Neuronal ramping activity during perceptual decisions. *A:* ramping activity during dot motion discrimination in monkey lateral intraparietal cortex (average from 54 neurons). Continuous and dotted lines indicate saccades into and out of neuronal response field, respectively. Motion strength indicates the percentage of coherently moving dots, and thus the facility to discriminate between their direction (out of two possible directions). The buildup is steeper with higher motion coherence but ends at same height at the time of choice. Stimulus marks onset of search array (target and distractors); choice marks saccade onset. [From Roitman and Shadlen (483).] *B:* ramping activity during visual search in a monkey frontal eye field neuron. Heavy and thin lines indicate target position inside and out of neuronal response field, respectively. Stimulus marks onset of search array (target and distractors); filled arrow marks saccade onset. [From Thompson et al. (588).] *C:* average ramping activity from 104 neurons in monkey lateral intraparietal cortex reflects subjective perception of dot motion direction rather than objective motion direction (12.8% coherently moving dots). Correct trials (full line) and error trials (dashed) differentiate between saccades into and out of neuronal response field (black and gray), irrespective of actual dot motion direction. [From Shadlen and Newsome (535).]

actions, and reward may constitute a basic neuronal mechanism for perceptual decisions, economic decisions, and internal movement generation.

During perceptual decisions, primary sensory systems code evidence about physical stimulus parameters irrespective of the subject's perception and decision (120), whereas activity in secondary sensory cortex and premotor cortex reflects the animal's subjective perception (120, 121). Ramping activity occurs over several seconds in neurons in parietal cortex, frontal cortex, and superior colliculus during the gradual accumulation of noisy evidence and reaches a criterion threshold at the moment of choice. Ramping activity is absent when sensory evidence is immediate and thus may not simply reflect neuronal processing (334). Ramping activity occurs during dot motion discrimination in prefrontal cortex, lateral intraparietal cortex (LIP), and superior colliculus **(FIGURE 38*A*)** (230, 274, 483, 535); length and brightness discrimination in superior colliculus (448, 450); and visual search and visual stimulus identification in frontal eye field neurons **(FIGURE 38*B*)** (49, 587, 588). Although typically tested with two choice options, ramping activity in LIP occurs also with four options (97). The slope of the ramp buildup to criterion threshold varies in LIP and superior colliculus with the speed and difficulty of the decision derived from the coherence of moving dots **(FIGURE 38*A*)** (483), differences in reaction time (97, 270, 448, 483), and differently rewarded options (494). The height of activity at choice threshold is the same irrespective of sensory coherence and motor response time in LIP **(FIGURE 38*A*)** (483). The statistical variance of neuronal activity in LIP increases gradually and declines earlier for one option compared with the other option immediately before the action (96). The probability of accurately predicting the animal's behavioral choice from neuronal activity in LIP increases during the ramp (535), suggesting that intraparietal activity is suitable for affecting subsequent neurons involved in expressing the decision through action. Accordingly, ramping activity in prefrontal neurons changes its direction while monkeys change their decision (269). Choice accuracy depends on the starting point of the neuronal ramp but not its threshold in LIP (199). The ramping activity in perceptual choice tasks does not correlate with objective sensory stimulus parameters but reflects the subjective perception expressed by the animal's choice, as shown with errors in direction judgement in LIP neurons **(FIGURE 38*C*)** (270, 535). Ramping activity may not always reflect motor processes, as it occurs in the frontal eye field even without action (587, 588) and in LIP correlates with motion strength evidence but not action (535).

During action decisions, even when straightforward sensory signals do not require lengthy accumulation of information, ramping activity in frontal eye fields builds up to a threshold for movement initiation before saccadic choices (198) and before correctly countermanded saccades **(FIGURE 39*A*)** (507). The rate of neuronal build-up determines saccadic reaction time **(FIGURE 39*B*)** (198, 508). Neuronal activity in cingulate motor area, frontal eye fields, and LIP ramps up before monkeys change movements to restore declining reward **(FIGURE 39, *C* AND *D*)** (101, 543). During delayed action instruction, neurons in premotor cortex

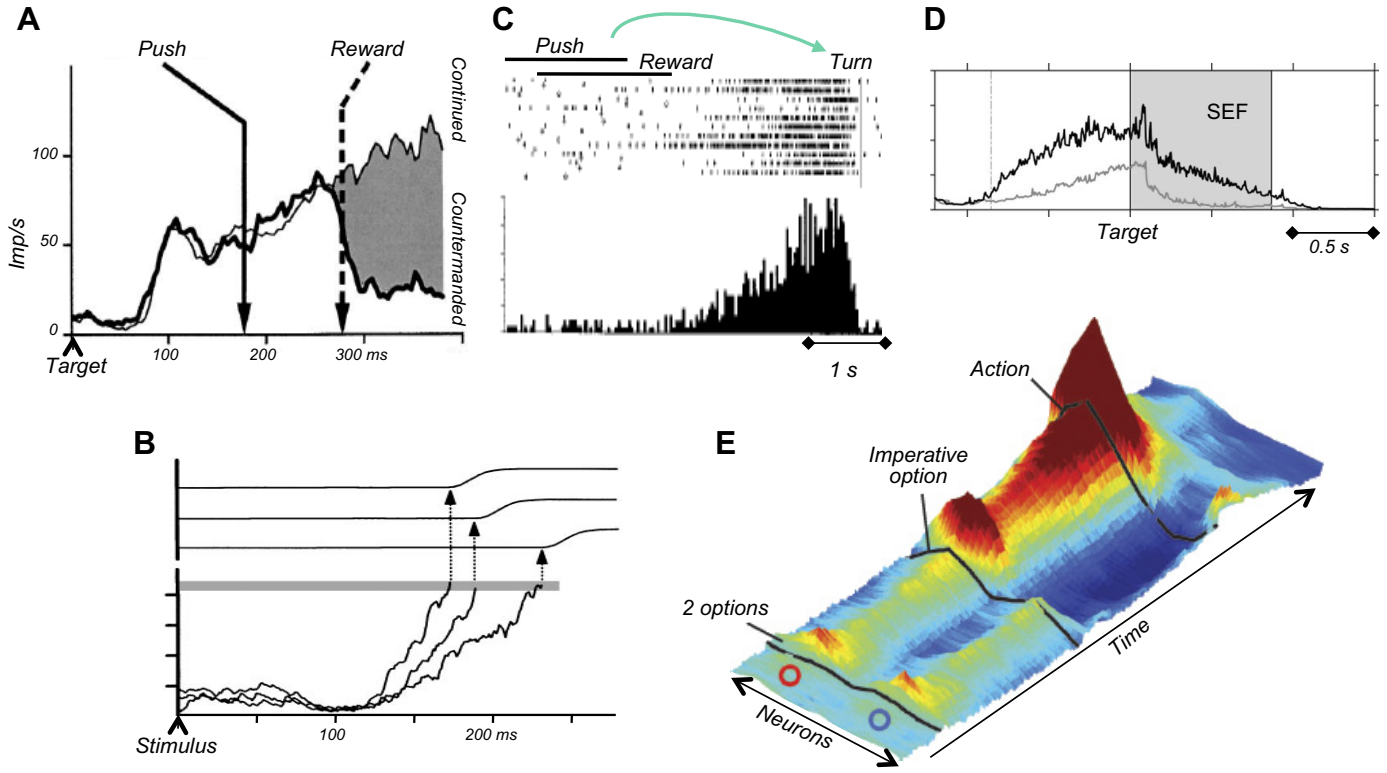


FIGURE 39. Neuronal ramping activity during action decisions. *A:* ramping activity during eye movement countermanding in a monkey frontal eye field neuron. Thin and heavy lines indicate continued and correctly stopped (countermanded) saccades, respectively. [From Schall et al. (507), with permission from Elsevier.] *B:* steeper ramping activity associated with earlier saccade choice in a monkey frontal eye field neuron. Neuronal activity has reached the same height at time of choice irrespective of slope. [From Schall and Thompson (508). Copyright 1999, Annual Reviews.] *C:* ramping activity during action decisions in monkey cingulate motor area. Upon decrease of reward amount, the monkey selects an alternate arm movement (turn) after the previous movement (push), thus restoring full reward amount. [From Shima and Tanji (543). Reprinted with permission from AAAS.] *D:* ramping activity preceding choice target reflects future choice in monkey supplementary eye field neuron. Black, free choice into neuronal response field; gray, opposite choice. [From Coe et al. (101).] *E:* differentiation of action coding in monkey premotor cortex. Neuronal population activity is initially segregated according to each action (red and blue circles) but becomes selective for final action after an imperative cue instructs the action. [From Cisek and Kalaska (99). Copyright 2010, Annual Reviews.]

show initial nondifferential activity that becomes stronger, selective, and ramps up for the action once the target is revealed **(FIGURE 39*E*)** (98). As these activities occur in motor areas, reaching the threshold may reflect the completed decision process or the initiation of action.

With economic decisions between differently rewarded options, neurons in LIP show ramping activity with noisy reward options (494). This activity likely reflects the accumulation of sensory evidence before the reward valuation rather than constituting the valuation process itself. However, LIP neurons show ramping activity also with immediately evident, non-noisy information about the reward options, such as differently colored lights at specific spatial positions **(FIGURE 40)** (571). The neuronal activity scales with the economic decision variable of fractional income and predicts the chosen action. Neurons in anterior cingulate cortex show ramping build-up towards a single threshold when animals leave a foraging patch with decaying reward value (209). Steeper ramp slopes reflect shorter reaction times for leaving the patch, as the overt choice is initiated when the common threshold is reached. These ramping activities cannot reflect noisy evidence but may constitute components of the neuronal decision process in the transition from sensory processing to reward valuation. They might also reflect associated processes such as stimulus or reward expectation or movement preparation. Nevertheless, ramping activity is not a necessary feature for economic decision activity in frontal cortex neurons (404), although its absence in a studied neuronal pool would not contradict ramping activity in other, unrecorded neurons, as cortical neurons show vastly inhomogeneous task relationships. In a probabilistic choice task with unambiguous visual symbols, parietal cortex activity reflects the economic decision variable of log likelihood ratio of probabilistic rewards (643). A potential ramp is due to averaging of asynchronous, nonramping activity. Without outright decisions, responses of amygdala and dopamine neurons reflect the

transition from initial sensory processing to reward valuation **(FIGURE 9)** (159, 422), which is considerably prolonged when the sensory evidence is noisy (389). Neurophysiological dopamine responses show a ramp with reward risk **(FIGURE 31, *C* AND *D*)** (161) but not with reward value. Striatal dopamine release ramps up before reward (237) which likely reflects presynaptic influences from reward expectation related cortical or amygdala inputs to striatum (40, 216, 544, 602) or from striatal neurons (18, 205, 215, 523) (Figures 32C and **37, *D* and *E***), but not a dopamine impulse ramp which does not occur with riskless rewards.

The morphology of ramping activity corresponds to the number of thresholds per neuron in the different decision models (56). The two-threshold drift-diffusion model assumes neuronal activity that ramps up differentially towards upper and lower thresholds for respective options **(FIGURE 36*A*)** (254, 302, 447, 565). Such activity is seen during perceptual decisions **(FIGURE 38, *A* AND *C*)** (97, 449, 483, 535) and economic decisions **(FIGURE 40)** (571). The observed gradual divergence of ramping activity in the same neuron may represent the output of an ongoing graded comparison with mutual inhibition in competing presynaptic neurons **(FIGURE 36*C*)**. The single-threshold type of decision model conceptualizes a gradual divergence of ramping activity in the same neuron. The ramp either reaches the threshold for a specific option or fades back to baseline for the alternative option **(FIGURES 36*B*)**. Such activity would also form the basis for race models in which activities from different pools of differentially active neurons race toward their individual thresholds (609, 614). Differential ramping activity is seen during the expectation of sensory events **(FIGURE 37*A*)** (6, 18, 65, 215, 248, 350, 526, 528), motor preparation **(FIGURE 37*B*)** (74, 98, 101, 109, 147, 184, 298, 299, 311, 337, 379, 380, 397, 467, 491, 493, 506, 526, 528, 636), reward expectation **(FIGURE 37, *D* AND *E*)** (18, 40, 205, 215, 216, 523, 544, 602), perceptual decisions **(FIGURE 38, *A* AND *B*)** (274, 483, 588), motor decisions **(FIGURE 39*A*)** (198, 507, 508, 587), and economic decisions **(FIGURE 40)** (571). The final step to the choice in these models may involve a graded or WTA competition between different neurons **(FIGURE 36*C*)**.

Taken together, neuronal ramping activities constitute key mechanisms in the identification of choice options from sensory evidence and may partly underlie their valuation. They reflect both the evidence that the decision maker cannot influence and the decision variables controlled by the decision maker. Ramping activities resemble statistical decision processes for accumulating stochastic information about sensory evidence (185) and rewards and follow major assumptions of decision models **(FIGURE 36, *A* AND *B*)**. The activities are widespread in the brain and reflect internal processes related to expectation of events, decisions between options, and preparation of action. As they occur also without choices, they may represent a general feature of neuronal physiology that is used for decision-making. Gradually increasing activity provides a focusing influence on downstream neurons through greater signal strength and higher signal-to-noise ratio, allowing better selection of behavioral actions **(FIGURE 36*C*)**.

4. Object value

Object value denotes the value of a specific reward object. It derives intuitively from nutrient or other survival values of rewards that are sought by humans and animals for survival. In human societies, object value is fundamental for economic exchanges of goods in which a quantity of one good with a specific value is traded against another good with a specific value. Object value is the basis for the marginal rate of substitution in economics which is the amount of *good A* a consumer is willing to give up to obtain one unit of *good B*, utility remaining constant across the exchange (341). To choose optimally, we assess the utility of each object separately, compare between them, and then select the best one. The value updating after each experienced outcome involves reinforcement learning (*Equations 10B, 10E, 26A*, and *26C*) with Pavlovian, habit, and goal-directed processes. Thus object value is defined as an input variable for competitive decision mechanisms (**FIGURES 35,** *top*, and **36*C***), in analogy to action value conceptualized in machine learning (575).

Object value has three crucial characteristics. First, it varies with the values of a reward object and, in principle, should be quantified as formal economic utility (*Equations 9–12*). Second, the value refers to one specific choice object. When choosing between several reward objects, such as a piece of meat and a loaf of bread, each object has its own and inde-

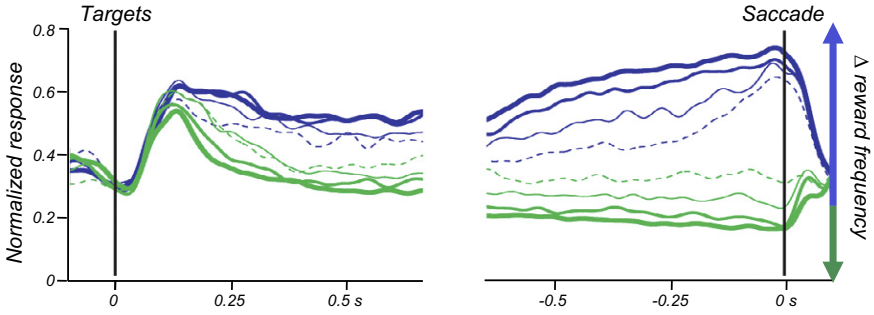


FIGURE 40. Neuronal ramping activity during economic decisions. Ramping activity in monkey lateral intraparietal cortex neurons during oculomotor matching choice between more and less frequently rewarded options (average from 43 neurons). Activity reflects the "fractional income" (ratio of chosen reward value to summed value of all options). Blue, saccade into neuronal response field; green, opposite movement. Fractional income varies between 0 (dotted lines) and 1.0 (think lines). [From Sugrue et al. (571). Reprinted with permission from AAAS.]

pendent value. "Objects" are solid or liquid rewards indicated by their appearance (color of fruit or fluid) and specific, intrinsically arbitrary stimuli associated with rewards. Different containers of the same kind of milk are examples of stimulus objects. Without being outright physical objects, events, situations and activities endowed with value have functions analogous to objects for economic decision mechanisms. The objects gain value through the animal's actual experience with the reward or from inferences based on non-Bayesian or Bayesian models of the world. Third, object value is independent of the actual choice of the object. It indicates the value of a good irrespective of whether we will imminently receive or consume it, thus complying with the notion of reward information, rather than explicit reward prediction. Economic exchanges are based on value comparisons irrespective of immediate consumption. A pound of meat has an object value of \$15 irrespective of whether I buy it now or not. Without this property, object values cannot serve as independent inputs to comparative decision mechanisms. Thus minimal tests for object value require two different reward objects, two values separately for each object, and independence from object choice.

Object value neurons process the value of one particular object irrespective of the animal's action. There would be grape juice neurons tracking the value of grape juice irrespective of the animal's choice, or banana neurons indicating how much the monkey will get if it were to climb the tree. The separate neurons tracking value for the specific objects constitute inputs for competitive decision models. In a binary decision, each input neuron, or pool of neurons, coding the value of its specific object competes with the input neurons coding the value of the other object. Hybrid subforms may show more pronounced object value coding if one particular object or action is chosen (combination with chosen object or chosen action coding defined below). To make meaningful comparisons, their signals need to vary on a common scale, both between the rewards ("common currency") and between the neurons. Neither sensory differences between rewards nor different coding ranges or slopes between neurons should affect object value signals. The model can be easily extended to three and more options as long as each neuron codes only the value of a single object. Thus object value coding on a common scale appears to constitute a suitable neuronal decision signal for competitive decision mechanisms.

Neurons tracking the value of specific reward objects irrespective of reward choice or reception are found in orbitofrontal cortex. They code selectively the value of a specific fruit juice irrespective of the monkey's choice and the required action [named "offer" value (405)] **(FIGURE 41, *A* AND *B*)** and thus do not predict the choice (404). These activities comply with the competitive formalism that requires independent coding of object value at the input **(FIGURE 36*C*)**. Orbitofrontal object value coding remains selective with three rewards (406), conforming coding specificity for a particular object. Impaired object-reward associations during choices in orbitofrontal lesioned humans and monkeys underline an orbitofrontal role in object value coding (84, 108).

In contrast, most conventional reward responses reflect explicit reward prediction or reception and thus do not have the necessary property of coding object value irrespective of the actual choice, without which they could not serve as independent inputs to competitive decision mechanisms. They reflect the actual, free or forced, behavioral choice and probably code forms of chosen value and reward prediction (see below).

5. Action value

All choices lead to movements. Whether they are ocular saccades, licking, reaching, locomotion, or speech, they all consist of muscle contractions performed to get the best value for the effort. Even verbally expressed choices, and speaking itself, involve muscle contractions (of the larynx, mouth, and hands). Machine learning and reinforcement learning, which are concerned with artificial agents moving about the world, use action value to model the agent's behavior for maximizing reward (575), like robots deciding which car part to weld first. Animals deciding between foraging patches are likely more interested in the value the action can bring them rather than considering each patch as an object. Each action leads to a reward value if it were chosen, thus the shorthand term of action value. Thus, for maximizing utility, an agent selects the action by which they achieve the highest estimated value. Pavlovian, habit, and goal-directed processes provide mechanisms for setting up and updating action value which, in principle, should be quantified as formal economic utility (*Equations 9–25*).

Actions are not only means for obtaining reward but may be themselves pleasurable and rewarding. The intrinsic reward may be added to the action value that usually refers only to the reward obtained from the action. The subjective nature of intrinsic action value emphasizes the need for quantitative behavioral assessments of often nonlinear and even nonmonotonic subjective value functions rather than inferring value from objective reward and action properties.

Like object value, action value is defined as input variable for competitive decision mechanisms (**FIGURES 35,** *top*, and **36*C***). Similar to object value, it requires three conditions. First, action value reflects reward value (amount, probability, EV, EU; *Equations 9–13*, *16*, and *21–25*). Second, value coding is specific for a particular action. The value is attached to an action and varies only for that action, such as an arm movement to the right or the left (spatial differential), an arm or an eye movement (effector differential), or going to work or vacation. Action value may arise from operant conditioning following the animal's actual experi-

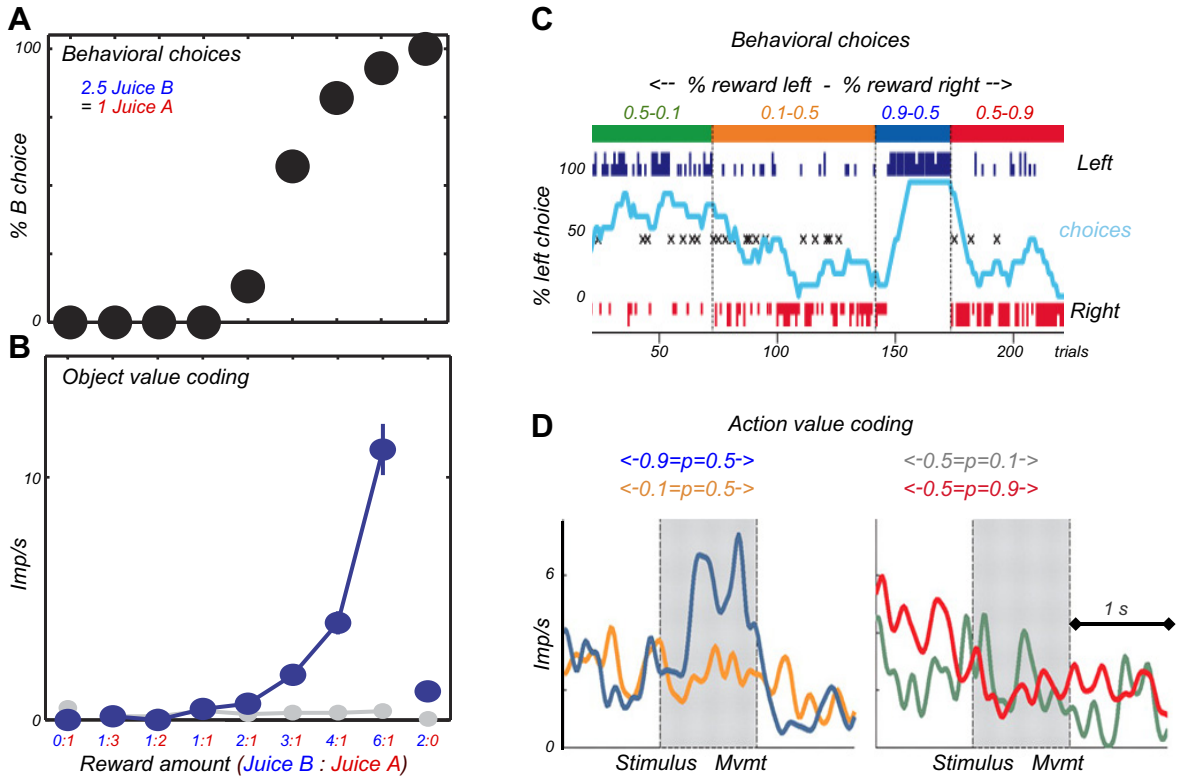


FIGURE 41. Neuronal coding of economic input decision variables for competitive mechanisms. *A:* behavioral choices between two juices that vary in magnitude. Increasing the amount of juice B (blue) increases the frequency of the monkey choosing that juice. The estimated indifference point (50% B choice) reveals that juice B is worth 0.4 units (1/2.5) of juice A. *B:* object value coding of single neuron in monkey orbitofrontal cortex during behavioral choices assessed in *A*. Activity increases with the amount of only juice B, not juice A, suggesting object value coding of reward B. [*A* and *B* from Padoa-Schioppa and Assad (405). Reprinted with permission from Nature Publishing Group.] *C:* behavioral choices between two actions that vary in reward probability. Blue and red ticks show actual choices of left and right targets, respectively. Light blue line shows running % left choices. *D:* action value coding of single neuron in monkey striatum. Activity increases with value (probability) for left action (*left panel*: blue vs. orange), but remains unchanged with value changes for right action (*right panel*), thus coding left action value. [*C* and *D* from Samejima et al. (500). Reprinted with permission from AAAS.]

ence with the reward or from inferences based on models of the world, e.g., Bayesian. Third, action value is independent of the actual action choice and the value obtained by that choice. Without this property, action values cannot serve as independent inputs to competitive decision mechanisms. Thus minimal tests for action value require two different actions, two different values for each action, and independence from choice.

Action value neurons process the value for one particular action irrespective of the animal's action. There would be left action value neurons tracking the value of a movement to the left irrespective of the animal's actual movement on that trial, and there would be separate right action value neurons. Or different neurons would respectively track the value of arm and eye movements irrespective of the animal's actual movement selection. Such separate action value tracking neurons are suitable inputs to competitive decision mechanisms that underlie utility maximization **(FIGURE 36*C*)**. Subforms may show preferential action value coding for particular chosen actions or objects. Action value signals need to vary on a common scale, both between the actions and between the neurons, to constitute comparable neuronal decision signals. Sensory and motor parameters of the action should not affect action value signals.

Neurons coding action values according to these definitions are found in monkey and rat dorsal and ventral striatum (slowly and tonically firing neurons), globus pallidus, dorsolateral prefrontal cortex, and anterior cingulate of monkeys and rats. These neurons code selectively the value of the left or right arm or eye movement irrespective of the animal's choice (**FIGURE 41, *C* AND D**) (245, 267, 275, 306, 500, 530). Action value neurons are more frequent in monkey striatum than dorsolateral prefrontal cortex (530). Action values in these studies are subjective and derived from behavioral models fitted to the choices (245, 267, 306, 500, 530) and from logistic regressions on the animal's choice frequencies (275). However, the required common scale coding is unknown. The anterior cingulate cortex is involved in action-reward mapping during choices, as lesions

there reduce preferences for more rewarded movements in humans and monkeys (84, 108).

The substantial action value coding in the striatum may reflect the pronounced motor function of this structure. Indeed, optogenetic stimulation in mice of dopamine D1 receptor expressing striatal neurons (direct pathway) induces within a few minutes behavioral preferences toward an operant touch or nose poke target associated with the stimulation, whereas stimulation of D2 receptor expressing striatal neurons (indirect pathway) induce behavioral dispreference (293). These data are compatible with the notion of striatal action value signals constituting inputs to a competitive decision mechanism located in the striatum or downstream structures.

In contrast, some reward signals reflect the explicitly predicted reward for specific actions while the actions are being planned or executed. These signals do not comply with action value coding, as they don't show the required independence from free or forced choice and action and thus are unsuitable inputs to competitive decision mechanisms (see below under "chosen value"). They code the chosen value during choices and reward prediction in choices or imperative trials. If the signals code in addition specifics of the actions they may reflect goal-directed mechanisms (see above and **FIGURE 20, *E* AND *F***).

6. Object value versus action value

The two decision variables derive from the intuition of objects having reward value that can be traded and require actions for obtaining them. The two variables may play different roles depending on the choice situation. Object value computations are likely to play a major role when values of objects change frequently and need to be tracked before choosing the required action. Once the object has been selected, its value can be directly mapped onto the action, without computing and comparing all action values. When I like to select a fruit at a new lunch place, I check the values of all available fruits (object value), and then I choose the highest valued fruit. Then I will know the action that gets me the chosen fruit (map the object to the action), without need to value each action and choose between their values. I can enjoy the feeling to soon get my beloved fruit, but that comes after the decision has been made and thus does not enter the competitive decision process. Indeed, monkey orbitofrontal neurons code the chosen value and map it onto action activity in dorsolateral prefrontal cortex to initiate the action ("good-to-action transformation," Ref. 81).

In contrast, when object values are rather stable but actions change frequently, action value would be the more appropriate decision variable. With frequent action changes, effort cost may change and should to be tracked continuously and subtracted from income value (*Equation 13*). At the time of choice, the different object values are mapped onto action values while subtracting effort cost of the currently required actions. In realistic choice situations, the progression from object value to action value may not be so stereotyped, nor would object value have primacy over action value or vice versa. Rather, object values, action values, object-action mapping, and motor plans would evolve with partial overlap depending on the choice situation. Neuronal processing of these variables and mechanisms, and the competition mechanisms at each stage, may involve separate neuronal populations in different brain structures. Indeed, orbitofrontal and anterior cingulate lesions induce differential deficits in object-reward and action-reward associations (84, 108).

7. Abstract decision

Once the comparison process has resulted in selection of the best reward object, the output of the selection process may be specified by an abstract decision signal that concerns only the decision and as such may precede the specification of the choice. Abstract decision signals code the choice in a binary, categorical manner and irrespective of value, object, or action.

Abstract decision signals are seen during perceptual decisions. They become apparent when decoupling the decision temporally or spatially from the subsequent reporting action. The signals occur without, separately from, or stronger than object or action signals in prefrontal, premotor, secondary somatosensory, and anterior cingulate cortex with vibrotactile stimulus comparisons **(FIGURE 42*A*)** (212, 312, 334, 340, 489, 490) and visual stimulus detection (360, 361). They are also seen as buildup activity in supplementary eye fields and superior colliculus during visual search (587) and random dot motion discrimination (230).

Abstract decision signals occur also in specific neurons in the amygdala that show graded reward value coding early during the trial and transition to later binary abstract decision coding **(FIGURE 42*B*)** (188). The decision signal from 1–50 pooled neurons predicts the upcoming choice correctly in 75–90% of trials, respectively, and remained predictive even with identical reward between the two options. These neurons are not activated during imperative forced-choice trials, nor do they code visual stimuli indicating the options or the subsequently chosen oculomotor actions. Abstract decision signals during economic choices in the amygdala extends its recently recognized reward function (422) to economic decision-making and challenge the view of a structure primarily known for fear responses.

8. Chosen value

Chosen value reflects the value of the specific object or action that the decision maker obtains from the choice. Its

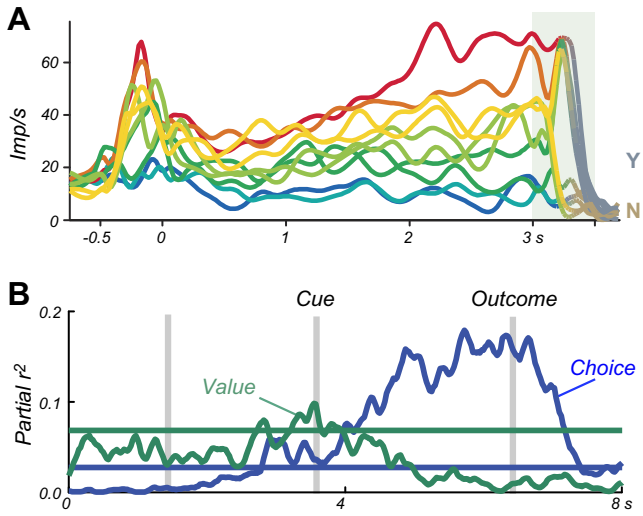


FIGURE 42. Abstract neuronal decision signals in monkeys. *A:* abstract perceptual decision coding in prefrontal cortex neuron. Initial graded frequency coding of vibrotactile stimulus (left peak) transitions to all-or-none coding of difference between two successive vibration frequencies. First vibratory stimulus begins at −0.5 s, second (comparison) vibratory stimulus begins at 3.0 s. Y and N refer to yes/no decision about the first stimulus frequency being higher than the second one. [From Machens et al. (334). Reprinted with permission from AAAS.] *B:* abstract economic decision coding during save-spend choices in single amygdala neuron. In each trial, the animal chose between saving reward with an interest or spending the saved reward. The neuron codes reward value early in the trial (derived from choice preferences, green) and later codes the actual save-spend choice (blue). Shown are time courses of the two most significant partial regression coefficients (r^2) from a multiple regression model that includes abstract save-spend choice, value (reward magnitude), spatial cue position, left-right action, and reaction time. [From Grabenhorst et al. (188).]

function is defined by competitive decision mechanisms (**FIGURES 35,** *middle*, and **36*C***). Whereas object value and action value are attached to an object or action irrespective of the choice, chosen value reflects the explicitly predicted or received reward value following the choice of that object or action. Thus chosen value reflects the output of the decision mechanism. For this important difference, chosen value should not be called object value or action value, even when it is attributed to a specific object or action. Chosen value is a decision variable, as it is controlled by the decision maker, who usually tries to maximize it (with the exceptions mentioned above). Chosen value, in principle, should be quantified as formal economic utility (*Equations 9–25*). Reinforcement learning employing Pavlovian, habit, and goal-directed processes provide mechanisms for setting up and updating chosen value.

Chosen value comes in two forms. Chosen object value reflects the value of a chosen reward object, like apple or orange, or the value of a specific stimulus predicting a particular reward. In contrast, chosen action value reflects the value of a chosen action. This distinction can be exemplified with particular reward magnitudes and probabilities. *Rewards A* and *B* are 1 ml and 2 ml of juice, respectively. *Reward A* occurs at the left in 20% of trials and at the right in 80%, whereas *reward B* occurs at the left in 60% and the right in 40% of trials. Then chosen object values are 1 ml and 2 ml for *rewards A* and *B*, respectively, whereas chosen action values are 1.4 ml for left (1 ml × 0.2 + 2 ml × 0.6) and 1.6 ml for right action (1 ml × 0.8 + 2 ml × 0.4). Most experiments do not specifically address this distinction by not using such discriminating values.

Neuronal chosen value signals reflect the value of an object or action irrespective of nonvalue attributes. These signals do not code the kind of reward (sensory aspects), sensory properties of predictive stimuli, spatial reward occurrence, stimulus positions, and required actions. In free choices between objects, specific neurons in orbitofrontal cortex code the chosen value of two or three different reward juices **(FIGURE 23, *A* AND *B*)** (405, 406). These activities reflect the estimated or experienced subjective economic value of the chosen juice, as assessed psychophysically by behavioral choices, and thus predict the choice (404). The signals scale in the same way across different rewards, thus coding value in common currency. Dopamine neurons code chosen value irrespective of action. Their value signal varies with the reward probability associated with the chosen stimulus (377). Stronger dopamine activations following conditioned stimuli predict the choice of the higher valued stimulus. Chosen value signals in the striatum show temporal discounting in choices between identical reward liquids delivered after different delays (79). Chosen value signals in anterior cingulate cortex code reward magnitude, probability, and effort, often in the same neurons (261). Chosen value signals in the supplementary eye field reflect the subjective value of objects, as estimated from certainty equivalents of behavioral choices between certain and risky outcomes (551). In imperative forced-choice trials, neurons in striatum and substantia nigra pars reticulata distinguish between stimuli predicting different reward magnitudes irrespective of sensory stimulus properties (107, 654).

In free choices between actions, some striatal neurons code chosen value without coding the action itself. The value signals vary with the reward probability associated with the chosen action and predict choice of the better action with higher activity during several task epochs (275, 306). Some striatal reward magnitude coding neurons do not distinguish between actions in imperative trials but reflect the reward value predicted from the choice (107). Chosen value signals in the striatum often occur after the overt behavioral choice, whereas action value signals typically appear before the choice (275, 306), suggesting that these chosen value signals represent the result of the decision process.

Unchosen value refers to the value of an unexperienced outcome. Whereas chosen value reflects the actual reward and is appropriate for experience-based decisions and reinforcement learning, unchosen value reflects some knowl-

edge about the world and might be more appropriate for model-based behavior. Animals benefit from the unchosen value information to improve reward gains (1, 208). Some neurons in dorsal and ventral striatum code unchosen value during intertemporal choices in monkeys (79) and during locomotor choices in rats (275). These neurons are distinct from the more frequent chosen value coding neurons in these structures. Some neurons code both chosen value and unchosen value with similar response slopes in anterior cingulate (208) and in orbitofrontal and dorsolateral prefrontal cortex (1). Thus some chosen value neurons code value irrespective of actual outcome.

Even before choosing, subjects may inspect the different options and process the anticipated value of each option. Humans show such inspectional eye movements whose frequencies and durations comply with evidence accumulation in diffusion decision models (291). It would be interesting to see whether neuronal value processing may initially reflect the value of the object or action being considered and subsequently shift to coding the chosen value. During action decisions, neurons in premotor cortex show initial activity for both actions but become selective for the final action after an imperative instruction (98, 99).

Hybrid forms of chosen value include chosen objects or actions. Corresponding neuronal signals reflect the value of the chosen reward together with sensory or motor information about the chosen object or action. If these activities occur well before the overt behavioral choice and increase towards a decision threshold, they may reflect the evidence accumulating in favor of the chosen option and thus play a role in the decision process. If the activities occur after the decision, they would constitute the outputs of the decision process and not drive the decision itself. They are occasionally referred to as object value or action value coding. However, according to the definition by competitive decision mechanisms **(FIGURE 36*C*)**, they constitute chosen value coding, as they reflect the reward value derived from the chosen objects or actions. Neuronal activity combining expected reward value conjointly with object information is only reported from imperative trials in frontal and temporal lobe neurons, as mentioned above (231, 319, 371, 394, 411), but cannot be labeled as "chosen" value due to the nonchoice task nature. Neurons coding chosen value conjointly with action are found with free choices in dorsal and ventral striatum (275, 306, 418, 479); globus pallidus (418); premotor, prefrontal, and anterior cingulate cortex (210, 532, 550, 620); and parietal cortex (433). The actions are eye or hand movements towards different spatial targets in monkeys (210, 306, 418, 433, 494, 550, 571, 620, 643) and nose pokes and whole left versus right body movements in rats (275, 479). In striatum, hybrid chosen value and action coding neurons are distinct from neurons coding only chosen value (306) or action value (306, 572). Conjoint coding exists also between unchosen value and action in orbitofrontal and dorsolateral prefrontal cortex (1). Comparable conjoint coding of value and action occurs also in nonchoice tasks (107, 135, 178, 205, 221, 243, 260, 282, 308, 313, 348, 354, 381, 427, 476, 502, 542, 606, 627), without qualifying for chosen value coding.

9. Relative value

The competitive decision mechanism for maximizing utility **(FIGURE 36*C*)** does not depend on actual values but involves comparison of the options relative to each other and is thus based on the difference or ratio of their values. Thus relative value is an important decision variable which, in principle, should be quantified as utility. Random walk-diffusion models capture the importance of relative value by assuming an early comparison with subsequent evidence growing toward specific thresholds. The mechanism for coding relative value may involve mutual inhibition between inputs that generates a graded difference between the inputs, amplified by the inhibition exerted by stronger inputs onto weaker inputs.

For binary decisions between *options 1* and *2*, the relative value can be stated using the final utility defined in *Equation 25*

$$\mathrm{REUsum} = \mathrm{EUsum1} - \mathrm{EUsum2} \qquad (27)$$

Instead of EUsum, any of the constituent utility functions defined by *Equations 9–13*, *16*, and *21–24* can be used for *Equation 27*. Most simply, REUsum may be derived from objective magnitudes $m1$ and $m2$

$$\mathrm{R}m = m1 - m2 \qquad (27A)$$

to obtain relative utility

$$\mathrm{Ru} = \mathrm{u1} - \mathrm{u2} \qquad (27B)$$

By applying a logarithmic utility function u = log (m), we obtain

$$\mathrm{Ru} = \log(m1) - \log(m2) \qquad (27C)$$

which is equivalent to

$$\mathrm{Ru} = \log(m1/m2) \qquad (27D)$$

As *Equations 8* and *9* show, reward value derives not only from reward magnitudes but also from reward probabilities. We replace $m1$ and $m2$ by probabilities $p1$ and $p2$

$$\mathrm{Ru} = \log(p1/p2) \qquad (27E)$$

Ru is a log odds ratio, also called log likelihood ratio (643). General simplifying formulations would replace objective m and p by their summed product of EV.

Another good measure of relative value is the ratio of objective values (magnitudes or probabilities)

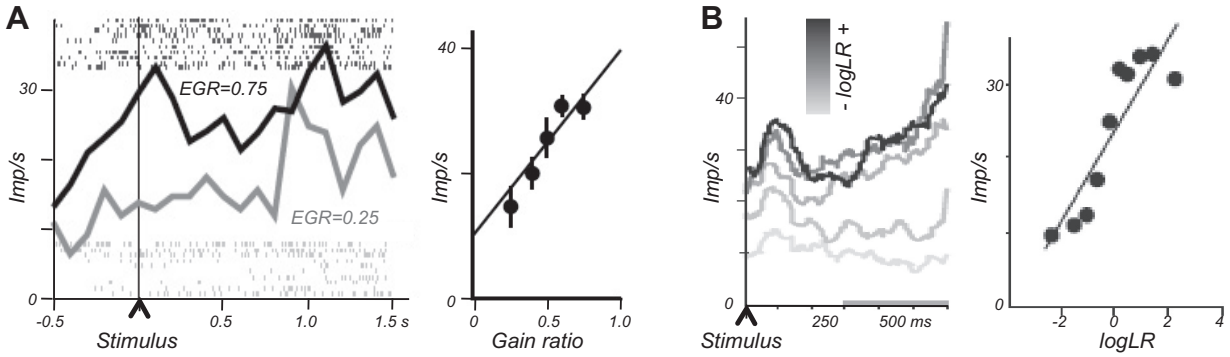


FIGURE 43. Relative value signals in monkey lateral intraparietal cortex. *A:* single neuron activity reflecting relative value during free oculomotor choice. *Left*: higher activity with higher "expected gain ratio" (EGR), lower activity with lower EGR. *Right*: linear regression on EGR. EGR denotes liquid reward volume of chosen option divided by summed magnitude of all options. [From Platt and Glimcher (433). Reprinted with permission of Nature Publishing Group.] *B:* single neuron responses increasing with the log likelihood ratio (logLR) of reward probabilities between chosen and unchosen options (choices into neuronal response field). [From Yang and Shadlen (643). Reprinted with permission from Nature Publishing Group.]

$$\mathrm{R}v = v1/v2 \qquad (27F)$$

The simple value ratio is often extended to include the summed value of all options in the denominator, called expected gain ratio (433) or fractional value (571)

$$\mathrm{RS}v = v1/\sum v_i;\ \text{for } i = 1, N;\ n = \text{number of options} \qquad (27G)$$

Reference value defined by *Equations 27-27G* applies to all forms of economic value, including object value, action value, and chosen value, and to all value functions defined by *Equations 8–13, 16*, and *21–24*. In monkeys, binary behavioral choices follow relative subjective value defined by the log odds ratio (*Equation 27E*) (643), reward magnitude ratio (*Equation 27F*) (306), or the ratio of objective chosen reward magnitude or probability over all options (*Equation 27G*) (433, 571). These measures are good predictors of choice preferences.

The difference value of *Equation 27* is analogous to reference-dependent utility (*Equation 22*), as both terms indicate relative value. Thus difference value and value adaptation serve the same function. They allow comparisons between existing options rather than requiring separate computations of the values of all options. Value ratios of the chosen option over the alternative (*Equation 27F*) or all options (*Equation 27G*) add adaptation to the spread of the reward probability distribution and correspond to divisive normalization in adaptive coding (332).

Some neurons in monkey and rat dorsal and ventral striatum, globus pallidus, and anterior cingulate cortex code difference subjective action value during free choices with eye or body movements, conforming to *Equation 27A* (79, 245, 306, 531, 572). Subjective value is derived from reinforcement models fitted to the choices or logistic regressions on choices. During temporal discounting, neurons code differences in subjective action value in dorsal striatum, and to some extent in ventral striatum (79). The discounted, subjective value is assessed from logistic regressions on intertemporal choices. Due to the relative nature of difference coding, neuronal activity varies positively with the value of one action or negatively with the value of the alternative action. True to the definition of action value, these relative value signals do not depend on actual choice. Neuronal difference signals exist also for chosen value in dorsal and ventral striatum, which take both chosen and unchosen values into account (79, 306). These activities increase with the value of one option and decrease with the alternative's value and, different from action value, depend on the actual choice. Neuronal activity in parietal association cortex reflects the reward ratio of the chosen option over all possible options (*Equation 27G*) **(FIGURE 43*A*)** (137, 433) and reaches a threshold at the time of choice **(FIGURE 40)** (494, 571). Similar parietal activity follows the increasing, logarithmic reward probability ratio between two options (*Equation 27E*) in the statistical sequential ratio test **(FIGURE 43*B*)** (643). Finally, difference signals for hybrid chosen value and oculomotor action exist in dorsal and ventral striatum (79, 306). Taken together, there are groups of distinct reward neurons coding basically every form of value decision variable in a relative way.

10. Chosen object and action

Although economic decision variables center around value, once the decision has been specified internally, the process may transfer to the overt behavioral choice without any longer involving value. Some neuronal signals in principal reward structures code the chosen object or the chosen action (frontal cortex, parietal cortex, and striatum) (**FIGURE 35,** *middle*).

Chosen object signals code the specific reward selected during the decision process or the specific visual stimulus pre-

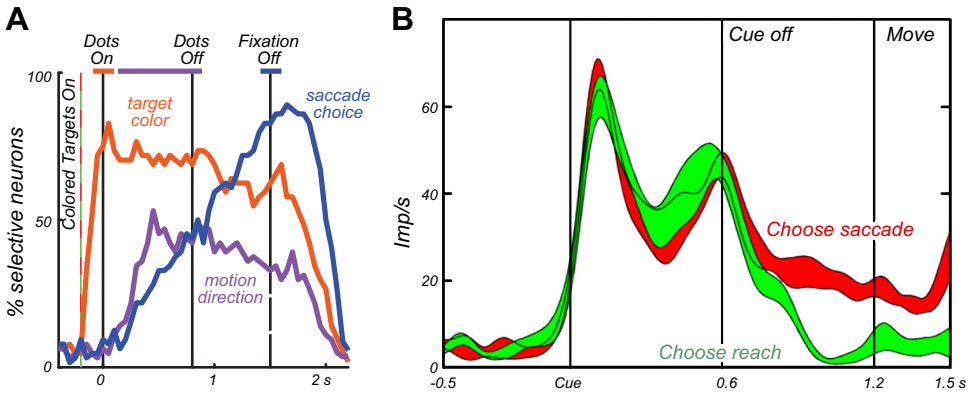


FIGURE 44. Chosen object and action signals in monkey lateral intraparietal cortex. *A:* coding of chosen object (orange color) or chosen action (blue) during perceptual decisions in population (averages from 23 neurons). In the conditional motor task, the animal saccades to a red or green dot (chosen object) depending on the direction of random dot motion. The dots are shown at different spatial positions and thus require different saccades (chosen action). Ordinate is % of significantly responding neurons for chosen object and chosen action (and dot motion direction, magenta). [From Bennur and Gold (38).] *B:* chosen action coding in single neuron. After a common onset, activity differentiates between effectors ~400 ms before action (saccade vs. reach). [From Cui and Andersen (109), with permission from Elsevier.]

dicting a particular reward. They serve to identify the chosen reward object while it is being acquired by an action. The signal does not code value and is not a simple visual or somatosensory response, as it may precede the appearance of the object. With economic decisions, chosen object signals in orbitofrontal cortex reflect different juices (405). With perceptual decisions, LIP neurons code the target color chosen for reporting the earlier, temporally distinct random dot motion direction (**FIGURE 44*A***, orange) (38). The signal is distinct from abstract decision and chosen action coded by other parietal neurons.

Chosen action signals code the specific action for acquiring the reward and may serve to prepare, initiate, and execute the necessary motor processes. These signals reflect the future action rather than reward value. They occur in cingulate motor area, anterior cingulate cortex, frontal and supplementary eye fields, and parietal cortex when monkeys change their actions after the reward for the current action declined (543), even before contacting a switch-over key (209) or before a chosen target occurs **(FIGURE 39*D*)** (101). The signals are absent or less pronounced in imperative forced-choice trials (101, 543). Neurons in LIP code the chosen saccadic action during probabilistic reward choice (433, 571), intertemporal choice (477), and dot motion discrimination (494), both during free choices and imperative forced-choice trials. In monkey and rat striatum, different neurons code chosen action from action value or chosen value during probability matching (306), locomotor choice (275), and intertemporal choice (79). Neurons in ventral striatum show much less chosen action signals than in dorsal caudate nucleus (79). Chosen action signals seem to begin earlier in striatal compared with parietal neurons and in both structures follow the subjective value signals by a few hundred milliseconds (79, 306, 477, 494). Some chosen action signals in parietal cortex are in addition modulated by chosen value (571). Thus chosen action signals occur in reward neurons close to behavioral output.

With perceptual decisions, parietal cortex neurons code the saccadic action for reporting random dot motion direction, distinct from abstract decision coding (**FIGURE 44*A***, blue) (38). Neurons in superior colliculus show chosen action signals superimposed on abstract decision signals (230).

With motor decisions, parietal cortex neurons show differential activity preceding freely chosen eye or arm movements **(FIGURE 44*B*)** (109, 509). During instructed action selection, separate neurons in premotor cortex initially code all options, but their activity differentiates and ramps up specifically for the arm movement imposed by a cue **(FIGURE 39*E*)** (98, 99). Although not reflecting free choice, the activity shows a nice transition from separate option coding to specific option coding as a general template for the progression of neuronal coding of economic decision variables.

11. Updating economic decision variables

Whereas perceptual decisions are based on sensory evidence, economic decisions require additional information about the value obtained by the choice. For meaningful decisions, this information needs to be predictive of the outcome. Choices without predictions amount to guessing. The value prediction does not derive from sensory organs but is learned and updated by error driven reinforcement learning. Its crucial error term is computed as the difference between the experienced reward and its prediction (*Equations 1, 2, 4, 5, 8A*–D, and *26-26C*).

The prediction entering the error computation derives from different sources (575). In model free reinforcement learn-

ing (460), the prediction derives only from direct experience with the values of the chosen objects or chosen actions in the past several trials. Other decision variables providing predictions, such as object value and action value, are only partly suitable for computing prediction errors because they are independent of the object or action being chosen and thus independent of the actually experienced reward. Their predictions are only useful for computing prediction errors when the objects or actions are chosen and their outcomes experienced, whereas unchosen rewards cannot enter the error computation. Thus the prediction in standard, model-free reinforcement learning derives from chosen value, which is a decision variable. In model-based reinforcement learning, the prediction incorporates information from models of the world and does not only reflect the recent experience with the outcome. The value prediction can be accurate even for objects and actions that have not recently been chosen and experienced. For example, given a model that assumes a constant sum of left and right action values, the left action value prediction can be accurate after only experiencing a right action and its outcome. Model-based reinforcement learning incorporates information beyond the experienced chosen value into the prediction and thus affects the error computation, whereas the reward is experienced as a matter of fact irrespective of model or not.

The type of prediction entering the error computation affects the form of the reward prediction error and thus the efficacy of value updating in specific learning scenarios (575). Three models are noteworthy in this respect. First, the Sarsa model applies to a range of laboratory tests in which the agent can experience all options during learning. Its prediction error incorporates the prediction derived from chosen value, which reflects the choice, but prediction from object value and action value is only valid when it derives from the actual choice. Thus chosen value neurons in striatum, orbitofrontal cortex, supplementary eye field, and anterior cingulate cortex (79, 208, 245, 261, 275, 306, 377, 405, 551) may serve as prediction inputs for the Sarsa model, and dopamine neurons code prediction errors in chosen value according to the Sarsa model (377). Second, Q learning directly approximates the optimal reward value within a given choice set. Its prediction error incorporates the prediction of the maximal attainable reward value, without taking the choice into account. Neurons in monkey anterior cingulate cortex code the prediction for the maximal probabilistic reward value that the animal usually but not always chooses (called task value) (12). Outright prediction error signals coding maximal value have not been described. Third, actor-critic models incorporate the sum of reward predictions from all options. Neurons in striatum, globus pallidus, and anterior cingulate cortex code the predicted sum (or mean with steeper slope) of reward values (79, 208, 245, 531). Taken together, the Sarsa reinforcement model has a correlate in the prediction error signals of dopamine neurons and other reward neurons that reflect predictions derived from chosen value. Future experiments might search for prediction error signals of the other reinforcement models in a wider range of learning situations.

Error signals suitable for updating decision variables are found in several reward structures. The bidirectional dopamine reward prediction error response (517) occurs with visual, auditory, and somatosensory stimuli (221, 365, 492) irrespective of action (521, 597, 618); integrates reward delay, reward risk, and different liquid and food rewards into a common subjective value signal (285, 301); incorporates model-based predictions (68, 382, 598); and reflects chosen value (377). Subgroups of non-dopamine neurons show bidirectional prediction error signals in lateral habenula, striatum, globus pallidus, amygdala, anterior cingulate cortex, and supplementary eye field (15, 36, 134, 227, 261, 275, 344, 531, 551). Some error signals in striatum code additional sensory and motor parameters (556). Neurons in cingulate cortex and supplementary eye field code bidirectional prediction errors from object value and action value to the extent of the object or action being chosen and leading to actual reward experience (chosen object value and chosen action value) (531, 550). Possible prediction error signals are also the separate, unidirectional positive and negative prediction error activations in anterior cingulate and supplementary eye field (246, 261, 551).

The updating function of the dopamine prediction error signal can be considered within the generic competitive model of decision-making **(FIGURE 45)**. Irrespective of the differences between formal decision models (55, 56, 441, 585, 623), the dopamine response would primarily update the inputs to the comparison process, namely, object value and action value that are coded during various trial epochs **(FIGURE 20, *A* AND *D*)**. The experienced and the predicted reward for the dopamine error computation likely derive from chosen value. The dopamine signal is rather similar among individual neurons and has heterogeneous effects on heterogeneous postsynaptic neurons. However, object value and action value are processed irrespective of the decision. They do not provide the necessary postsynaptic specificity to make the dopamine influence selective, which would indiscriminately update all object values and action values, including those not even experienced. However, as primary and higher order rewards occur only after an event, neuronal updating involves stimulus traces that maintain a label of the engaged synapses beyond the event and make them eligible for modification (574). Thus each object value and action value signal may leave a distinct eligibility trace that decays rapidly unless being stabilized by neuronal inputs **(FIGURE 14, *B* AND *C*)**. Neuronal activity from the specific chosen object/chosen action signal may provide selective inputs to object value/action value neurons for the same object/action for stabilizing their eligibility traces (**FIGURE 45,** *right*), whereas the lack of activity from the unchosen option would not prevent the decay (*left*). Thus the

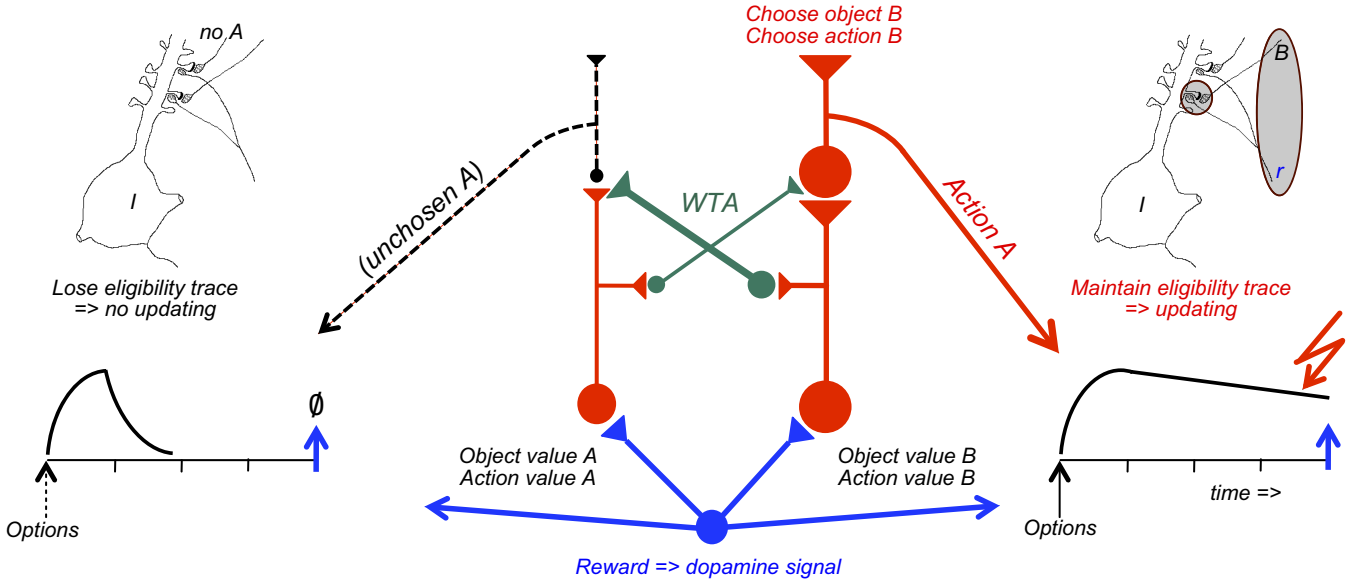


FIGURE 45. Hypothetical mechanism of dopamine updating of decision variables. Within a competitive decision mechanism, the global dopamine reward prediction error signal would act indiscriminately on postsynaptic neurons and changes their synaptic input efficacy by influencing stimulus eligibility traces. It affects only neurons coding object value or action value of the option chosen, as only their eligibility traces are being stabilized and maintained by input from neurons activated by the chosen object or chosen action (*right*), but not affect neurons whose initial eligibility traces are lost due to lack of stabilizing input from neurons not being activated by the unchosen object or unchosen action (*left*). This selective effect requires specific connections from chosen object/chosen action neurons to object value/action value neurons for the same object/action. The prediction error conveyed by the dopamine neurons derives from chosen value (experienced minus predicted reward). Gray zones at *top right* indicate common activations. The weight of dots and lines in the circuit model indicates level of neuronal activity. Dotted lines indicate inputs from unchanged neuronal activities. Crossed green connections are inhibitory; WTA, winner-take-all selection. Scheme developed together with Fabian Grabenhorst.

maintained eligibility trace reflects the output of the decision process and provides the necessary specific label for dopamine influences to act upon. As a result, the indiscriminate dopamine signal would selectively update the object value and/or action value of the chosen option, whereas object value and/or action value of the unchosen option remain unaltered and need to wait for their next selection before being updated (except in model-based reinforcement learning, see above). Thus the influences of chosen object and chosen action on the eligibility trace provide the necessary selectivity for dopamine action. As an example **(FIGURE 45)**, choice of a movement to the right and experience of the corresponding reward would lead to updating of right action value by a prediction error signal. If the right choice resulted in a higher than predicted value and thus a positive prediction error, right action value would be increased; however, if the value is lower than predicted and elicits a negative prediction error, right action value would be decreased. If the right action is not chosen, its value cannot be experienced and thus will not be updated (or might decay depending on the algorithm implemented). Taken together, within a simple comparator model of decision-making, the global influence of the dopamine error signal derived from chosen value may serve to update input decision variables for specific options via eligibility traces selectively stabilized by activity from the chosen option.

Neuronal correlates for this hypothetized dopamine influence consist of the global projection of this reward signal to most or all neurons in striatum and frontal cortex and to many neurons in amygdala, thalamus, and other structures **(FIGURE 17*A*)**. The released dopamine influences the plasticity between conjointly active presynaptic and postsynaptic neurons in a three factor Hebbian learning scheme **(FIGURE 17, *B* AND *C*)** involving striatum, frontal cortex, and other structures in which appropriate dopamine dependent plasticity is found (82, 194, 265, 294, 400, 401, 424, 461, 495, 540, 583, 642, 648). In this way, the dopamine signal could affect synaptic transmission onto striatal and cortical neurons coding object value and action value. These effects may be differential within striatal regions, as action value neurons are more abundant in dorsal compared with ventral striatum (245), whereas ventral striatum neurons are more involved in reward monitoring unrelated to action coding (245, 523). The dopamine plasticity effects may also be differential for striatal neuron types, as optogenetic stimulation of neurons expressing D1 or D2 dopamine receptors induces learning of behavioral preferences and dispreferences, respectively, possibly by updating action values **(FIGURE 19*F*)** (293). These neuronal data support the scheme of dopamine updating of decision variables **(FIGURE 45)**.

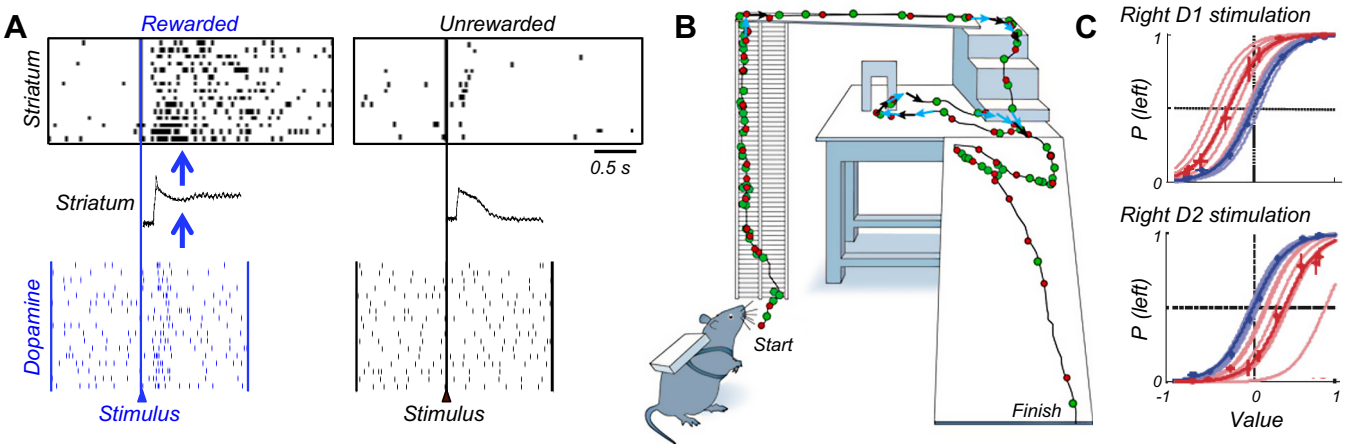


FIGURE 46. Immediate dopamine influences. *A:* hypothetical mechanism of influence of dopamine reward signal on striatal reward response. *Bottom*: a rewarded, but not unrewarded, stimulus differentially activates a dopamine neuron. [From Tobler et al. (598).] *Middle*: possible dopamine influence on striatal synaptic potentials: dopamine D1 receptor agonist prolongs the depolarization of a striatal neuron. [From Hernández-Lopez et al. (213).] *Top*: as a result of the dopamine influence on striatal depolarization, the striatal neuron responds stronger to a rewarded compared with an unrewarded stimulus. [From Hollerman et al. (222).] *B:* influence of dopamine reward signal on behavioral navigation. Electrical stimulation to the left and right somatosensory cortex of rats provides cues for turning; electrical stimulation of the medial forebrain bundle containing dopamine axons induces forward locomotion. The combined stimulation is able to guide a rat through a three-dimensional obstacle course, including an unnatural open field ramp. Colored dots indicate stimulation of forebrain bundle and somatosensory cortex. [From Talwar et al. (580). Reprinted with permission from Nature Publishing Group.] *C:* influence of dopamine reward signal on behavioral choices. Unilateral optogenetic stimulation of mouse striatal neurons expressing dopamine D1 or D2 receptors induces immediate differential choice biases (left bias for right D1 stimulation, *top*; and right bias for right D2 stimulation, *bottom*). [From Tai et al. (576). Reprinted with permission from Nature Publishing Group.]

Updating with non-dopamine error signals may be less global and more selective. These neurons likely exert focused influences on specific postsynaptic neurons through selective anatomical projections. Their error signals are more specific for the type of information they provide and for the type of postsynaptic neuron they influence. Thus they are likely to be more specific for the objects and actions and the type of neuronal value signals they update.

12. Immediate dopamine influences

In addition to affecting synaptic plasticity and learning, phasic dopamine signals exert immediate influences on postsynaptic processing and behavioral performance and choices. The anatomical basis may be the triad arrangement of dopamine synapses on glutamatergic inputs to dendritic spines of striatal and cortical neurons **(FIGURE 17*B*)** (171, 186) and convergence in striatum and globus pallidus (416, 430). The long debated focusing effect assumes that dopamine signals suppress the effects of weaker inputs to the striatum while allowing stronger ones to pass through to the basal ganglia outputs (70, 368, 594). Thus phasic dopamine signals may affect the function of existing anatomical convergence and modulate the use of neuronal connections. At the cellular level, the phasic dopamine signal has excitatory action on D1 receptor expressing striatal neurons of the direct pathway by prolonging transitions to membrane up states (depolarization) (213). In contrast, dopamine action on D2 receptor expressing striatal neurons of the indirect pathway reduces membrane up states and prolongs membrane down states (hyperpolarization) (214). In a hypothetical model of such action, the dopamine enhancement of cortical input efficacy may induce reward related activity in striatal neurons **(FIGURE 46*A*)**. This influence may affect reward processing in all forms of striatal task related activity **(FIGURE 20, *A* AND *D*)**. Impairments of dopamine transmission reveal the immediate dopamine influences. Striatal dopamine depletion reduces learned neuronal responses in striatum (13), alterations of D1 receptor stimulation affect mnemonic behavior and neuronal activity in frontal cortex (505, 616), and reduction of dopamine bursting activity via NMDA receptor knockout prolongs reaction time (650). These deficits relate well to the immediate behavioral effects of phasic dopamine stimulation. Optogenetic activation of mouse dopamine neurons, in addition to providing reinforcement, elicits immediate behavioral actions, including contralateral rotation and locomotion (271). Pairing electrical stimulation of the medial forebrain bundle containing dopamine axons with lateralized stimulation of somatosensory cortex induces forward locomotion in rats and directs the animals in sophisticated and even unnatural spatial navigational tasks **(FIGURE 46*B*)** (580). The forebrain bundle stimulation likely serves as immediate reward, whereas the cortical stimulation provides the necessary spatial information. Corresponding to the action-inducing effects of dopamine stimulation, indirect inhibition of VTA dopamine

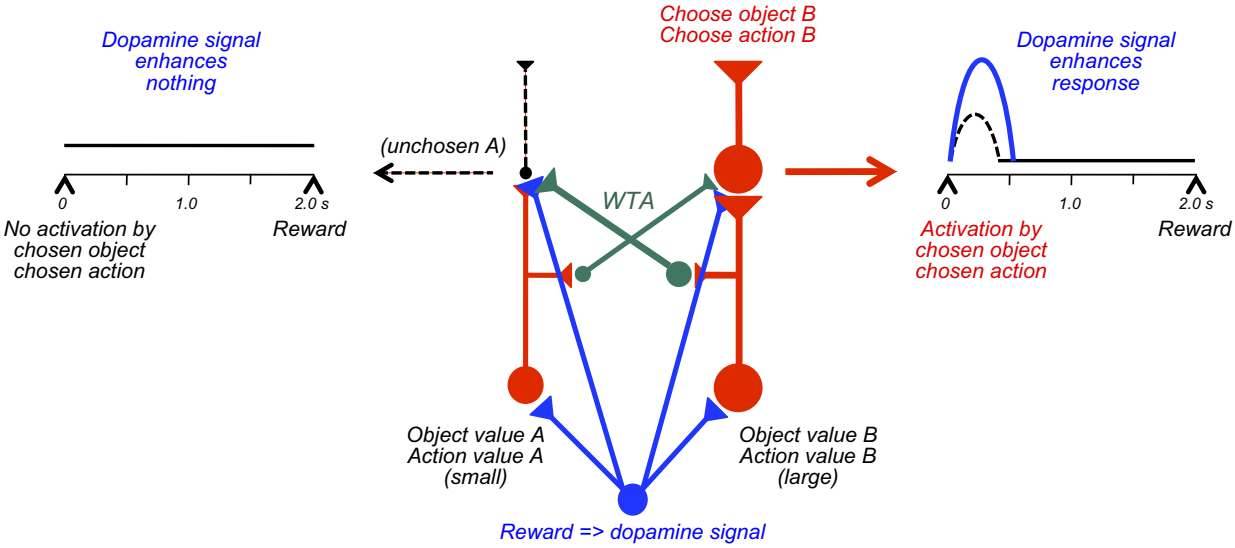


FIGURE 47. Hypothetical mechanism of immediate dopamine influence on neuronal coding of decision variables. For the basic competitive decision mechanism, the dopamine signal arises from a temporal difference (TD) prediction error in chosen value at the time of the decision. The dopamine signal boosts differences in activity between the two options reflecting object value or action value (*bottom*). Alternatively, and more effectively, the signal may enhance activities representing chosen object or chosen action selected by the winner-take-all (WTA) mechanism (*top right*), while leaving nonexisting activity unchanged (*top left*). The weight of dots and lines in the circuit model indicates level of neuronal activity. Dotted lines indicate inputs from unchanged neuronal activities. Crossed green connections are inhibitory.

neurons via optogenetic activation of local, presynaptic GABA neurons reduces reward consumption (613). At the postsynaptic level, optogenetic stimulation of dopamine D1 receptor expressing striatal neurons increases behavioral choices toward contralateral nose poke targets, whereas stimulation of D2 receptor expressing striatal neurons increases ipsilateral choices (or contralateral dispreferences) **(FIGURE 46*C*)** (576), suggesting an immediate, differential effect on reward value for the two actions.

Whereas these effects are straightforward, stimulating a couple synapses upstream of dopamine neurons induces surprisingly complex effects. Optogenetic activation of tegmental inputs to dopamine neurons projecting to nucleus accumbens elicits place preference, whereas stimulation of habenula inputs to dopamine neurons projecting to frontal cortex induces place dispreference (303). Whereas place preference signals reward, dispreference reflects either aversion or reduced reward. The dispreference might be due to monosynaptic excitation of aversive coding dopamine neurons, as suggested by FOS gene expression and excitatory currents (EPSCs) in dopamine neurons and by blunting of the dispreference with prefrontal D1 antagonists (303). However, dopamine neurons are not activated by aversive stimulus components (157, 160), FOS activation is slow and may derive from rebound activation following dopamine depression (158, 566), EPSCs do not necessarily induce action potentials required for dopamine release in target areas, and the optogenetic stimulation of presumed dopamine neurons might have included non-dopamine GABA neurons in used TH:Cre mice (304). Alternatively, dispreference might reflect disynaptic inhibition of dopamine neurons via the inhibitory rostromedial reticular nucleus (613). The monosynaptic habenula-dopamine projection is weak (625), and electrophysiology reports depression-activation sequence of VTA and nigral dopamine impulse activity following habenula stimulation (95, 250, 344, 566) that likely involves the inhibitory rostromedial reticular nucleus (226, 249). Furthermore, habenula glutamate receptor blockade increases striatal and prefrontal dopamine concentrations, probably reflecting blockade of the inhibitory habenula-dopamine projection (310). The prefrontal D1 antagonists likely blunt all prefrontal dopamine effects irrespective of excitation or inhibition. Thus the place dispreference with habenula stimulation unlikely reflects monosynaptic activation of aversive coding dopamine neurons and rather results from inhibiting dopamine neurons via the reticular nucleus. Indeed, a recent study reports inhibition of dopamine neurons and behavioral dispreference by electrical habenula stimulation. Similar place dispreference is elicited by activating midbrain GABA neurons that inhibit dopamine neurons (613) or by directly inhibiting dopamine neurons (244, 582). Thus the approach and dispreference elicited by differential input stimulation are likely the results of transsynaptically induced increased and reduced dopamine reward signals, respectively.

The immediate behavioral dopamine effects may be captured by a simplifying account that incorporates the essentials of immediate dopamine influences on behavioral choices into a generic competitive decision model **(FIGURE 47)**. To have an immediate effect, the dopamine signal

should reflect the prediction error at the time of the decision preceding the final reward. This error is the TD reward prediction error between the current prediction reflecting the chosen value minus the preceding reward prediction. The signal may influence neuronal activity for two kinds of decision variables. A dopamine-focusing effect (70, 368, 594) on predecisional object value or action value signals would enhance the value difference between the two options in a graded manner. The indiscriminate dopamine signal would affect all active synapses on neurons coding by definition object value and action value irrespective of the decision. Such dopamine influences on action value coding were assumed to explain the differential effects of optogenetic stimulation of D1 and D2 containing striatal neurons on behavioral choices (576). A more divisive dopamine influence would be achieved by affecting neurons that code the postdecisional variables of chosen object or chosen action before the overt behavioral choice (**FIGURE 47**, *right*). Activity in these neurons is differential due to selection by the WTA decision mechanism. The nonexisting activity from the unchosen option would not be enhanced by the dopamine signal (left). Thus the WTA mechanism would provide selectivity for the indiscriminate dopamine influence. Taken together, the basic comparator model of decision-making provides an architecture possibly underlying the immediate influence of the phasic dopamine signal on decision processes. This hypothesis is intentionally crude and provides only a general account, without addressing details of implementation by different types of dopamine innervated neurons in striatum, frontal cortex, and amygdala.

13. Formal neuronal decision models

The remarkable correspondence of neuronal signals to the key features of decision models encourages the development of biologically inspired neuronal models of decision-making. In many of such models, the economic decision variables of individual choice options are processed in separate pools of spiking neurons. The ramping activity is generated from transient inputs by recurrent synaptic excitation involving AMPA or NMDA receptors to induce self-sustained attractor states and integrate sensory or value information. The competition between choice options is mediated by GABA neurons exerting mutual, lateral, feedforward, or feedback inhibition between the neuronal pools for the choice options. The WTA mechanism combines the inhibition with a thresholding mechanism that leads to acceleration of ramping activity or to all-or-nothing selection between signals. The winning suprathreshold activity then propagates to pools of output neurons that mediate the behavioral action.

For neuronal decision models involving the cerebral cortex, key features are the discovery of explicit neuronal decision signals and the local synaptic connectivity. A generic model for perceptual discrimination of random dot motion direction in LIP neurons uses a difference signal derived from pools of specific direction-sensitive input neurons from middle temporal (MT) cortex (535). For example, larger activity in rightward compared with leftward motion coding MT neurons results in an LIP difference signal for rightward motion. The activity begins with appearance of the moving dots and accumulates over time with the continuing incoming noisy evidence about dot motion direction until it reaches one of the two decision thresholds. This model conforms to a diffusion model with continuous comparison between options and activity diverging towards one of several thresholds (55) and fits well the experimental data (535). Through its generality, such a model would also fit economic decision processes during which LIP neurons code the log likelihood ratio decision variable in a sequential probability ratio test (643), which constitutes an efficient decision mechanism (619).

A cortex model of two-stage somatosensory frequency discrimination uses two mutually inhibitory pools of 250 noisy spiking neurons that represent the two choice options (334). The model neurons show sustained activity that scales with the frequency of the first, transient vibratory stimulus until the comparison stimulus occurs. The subsequent activity conforms to an abstract decision signal that does not vary with individual stimulus frequencies. The sustained activity and its transition from mnemonic coding to decision signal resemble closely the activity of prefrontal and premotor cortex neurons of monkeys in this task (65, 212, 488, 489). The model achieves categorical decisions by discriminating between vibratory stimuli irrespective of their actual frequencies and thus supports the function of the experimentally observed neuronal signals in perceptual decisions.

More elaborate cortical neuronal decision models employ architectures comprising frontal and parietal cortex, basal ganglia, and superior colliculus **(FIGURE 48*A*)** (323, 622, 623). Transient inputs induce persistent activity with slow temporal integration via recurrent cortical NMDA and AMPA mediated excitation. Mutual lateral inhibition between pools of 240 irregularly spiking neurons mediates the competition between the stochastic choice options. By excitatory attractor dynamics and WTA competition throughout decision formation, the ramping activities in the neuronal pools representing the choice options become increasingly more separated in their race toward specific decision thresholds. The ramping activities reflect evidence accumulation or decision variables such as event likelihood ratio for perceptual decisions or object value, action value, and derivatives for economic choices. The cortical activity exceeding the competition threshold is detected by the superior colliculus via direct or double inhibitory influences through caudate nucleus and pars reticulata of substantia nigra and then propagated to the oculomotor system. The modeled neuronal activities replicate well the experimentally observed

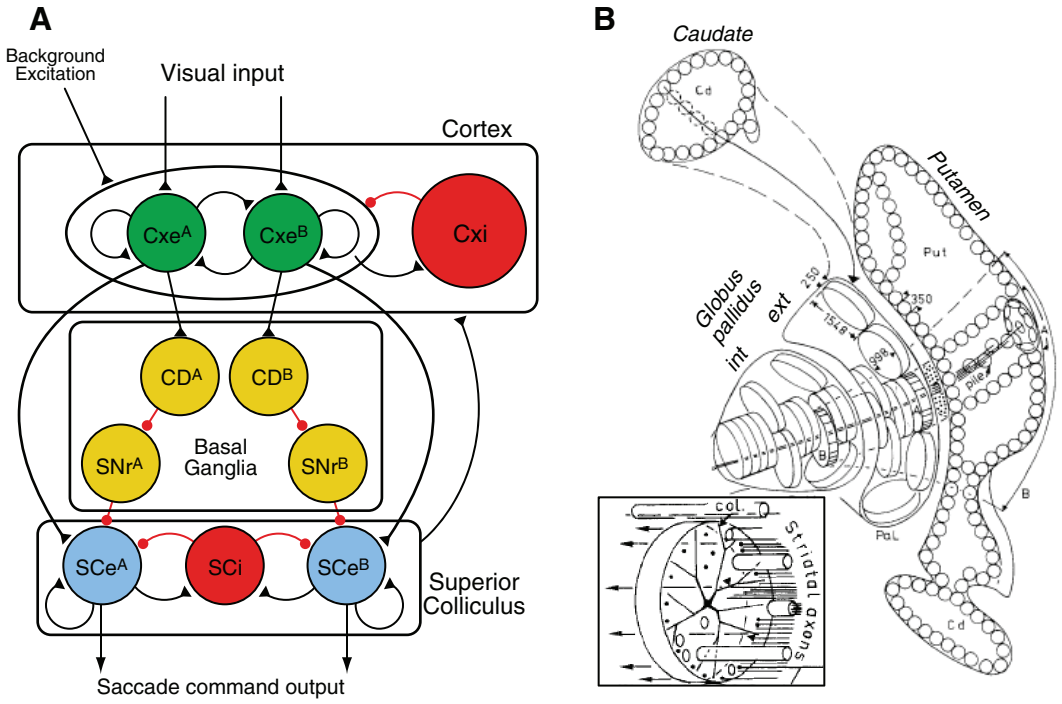


FIGURE 48. An example economic decision model and key anatomical foundations. *A:* architecture of a sensory or value decision model involving the cerebral cortex and basal ganglia. Recurrent excitations induce ramping activity in cortex (input to basal ganglia) and superior colliculus (output of basal ganglia), whereas lateral inhibition between neuronal pools representing the two choice options mediate competitive winner-take-all (WTA) selection of the better option at the input (cortex) and induce a corresponding ocular saccade at the output (superior colliculus). [From Wang (623), with permission from Elsevier.] *B:* anatomical convergence of neurons from striatum (caudate and putamen) onto dendritic disks of neurons in globus pallidus (internal and external segments) in monkey. The dendritic disks represent an abstraction of the wide dendritic arborizations that are oriented orthogonally to axons traversing the globus pallidus from the striatum. *Inset* shows several axons from striatal neurons traversing and contacting the pallidal dendritic disks. These parallel striatal axons induce substantial anatomical striatal-pallidal convergence by contacting the dendrites in the disks. [From Percheron et al. (430). Copyright 1984 John Wiley and Sons.]

ramping in frontal and parietal neurons described above. Dopamine-mediated three-factor Hebbian reinforcement learning may set the decision threshold in caudate nucleus or cortex (554). Thus cortical WTA determines the ramping, superior colliculus detects threshold crossing, and dopamine neurons set the threshold.

For neuronal decision models involving the basal ganglia, key features are the well-organized corticostriatal projections, internal connections, and parallel loops with cortex (7, 141, 164, 196, 197, 252); the synaptic convergence through the component structures **(FIGURE 48*B*)** (416, 430); the separate striatal outputs with sequential inhibitory neurons (93, 127); and the dopamine-mediated plasticity of corticostriatal connections (424, 517, 540). The basal ganglia may be involved in two distinguishable aspects of decision-making, namely, the direct selection of options and the updating of option values through reinforcement learning (57).

Option selection in the basal ganglia employs the dopamine focusing effect (70, 368, 594) and anatomical convergence in striatum and globus pallidus (416, 430) **(FIGURE 47)**. The models employ physiological activities and known input-output relationships of basal ganglia nuclei to demonstrate the feasibility of movement selection (72). By implementing the efficient sequential probability ratio test, the scaled race model shows that the subthalamic nucleus modulates the decision process proportionally to the conflict between alternatives (57). It outperforms unscaled race models. Particularly good fits are obtained with a diffusion model of movement selection (449). The basal ganglia decision models are supported by experimental data not used for their construction and also elucidate economic decision mechanisms, including the differential effects of optogenetic stimulation of dopamine D1 and D2 receptor expressing striatal neurons **(FIGURE 46C)** (576). The opposing functions of the direct and indirect striatal output pathways are modeled by a race between movement and countermanding signals that reflect the relative timing of the two signals as they race towards a threshold (511). The signals fit experimentally observed activities in rat substantia nigra neurons. These studies provide key support for important basal ganglia contributions to motor selection and economic decisions.

Updating of object value, action value, and their derivatives by the basal ganglia employs dopamine prediction error signals (517) **(FIGURE 45)**. Data analyses routinely employ the modeled subjective values of choice options as regressors for neuronal value coding and action value updating in the striatum (245, 267, 275, 306, 500, 530). Reinforcement models distinguishing between direct and indirect striatal pathways are based on in vitro work (424, 540) and assume that dopamine prediction error signals have differential plasticity effects on D1 versus D2 receptor expressing striatal neurons (72) and induce the learning of respective choice preferences and dispreferences (169). These results from formal models closely reflecting physiological data confirm the capacity of dopamine reinforcement signals for learning and updating of economic decision variables.

E. Reward and Free Will

Current notions about free will range from extreme determinism, in which every event of every individual at every moment is predetermined in advance by the mechanics of the world, to the capricious, unrestrained free will that puts intention and responsibility at each individuals' complete disposition and initiative. The truth lies probably somewhere in between. There are likely neuronal mechanisms involving stochasticity that are constrained by external influences and generate choice options within a limited range. Deliberate actions might originate from these neuronal mechanisms and operate within these constraints. Constraints are provided by reward value and all factors influencing it, including effort cost, delay, risk, and references, that affect brain processes underlying decision-making. Without trying to provide a coherent view, I will sketch a few thoughts along these lines.

1. Origin of free will (the impossible question)

Assuming that free will exists at all, we simply do not know how it may originate, and we have only faint ideas how brain processes might generate it. We argue even whether free will originates in the brain. For Kant, free will lies above all activity, irrespective of physical, biological, psychological, and social constraints. For him, only God has free will, as s/he has created everything. But admitting that we do not know anything is the easy answer. We can try to reduce this impossible question to tractable issues by trying to understand the mechanisms determining free will.

Free will has several necessary components. First, free will depends on the ability to make choices, to do what we want to do (219). Second, free will implies a voluntary action that is self-directed and arises on self-command. Free will does not apply to vegetative reactions that I cannot control, like heartbeat, intestinal contractions, or hormone release, nor to probably most Pavlovian reactions and many habits. Third, the word *will* indicates a goal that we want to achieve. Thus free will would be instrumental in generating voluntary, goal-directed behavior. Fourth, free will involves conscious awareness about the goal directedness. We think we have free will because we know explicitly that we are working towards something we want. And we know that we want this something, however diffuse this may be in exploratory behavior. We feel the freedom to choose when to act and which of several options to select. This feeling is subjective, private, and unobservable by others. It is difficult to measure objectively, and it may not even be real but an illusion (634), nobody knows for sure.

The fact that free will has emerged in evolution suggests a beneficial function for the competitive survival and reproduction of individuals in their function as gene carriers. As much as the idea of a conscious illusion of free will may be disappointing, evolutionary biology teaches us that a structure or function may enhance the chance for survival without necessarily evolving to perfection. Thus it may not matter whether conscious free will is an illusion or real, as long as it is helpful for keeping the genes in the pool. Thus free will is a psychological tool for gene propagation.

What we do know is that we need rewards for survival (essential substances contained in food and drinks) and reproduction (sexual rewards). It would be advantageous to seek these hidden rewards actively rather than waiting for them to appear out of the blue. That is why we can move. Intentions and free will would be a great help to seek and explore rewards actively and deliberately, rather than running after them using deterministic programs. Without the true or false belief of free will, we may have only limited initiative to find hidden rewards. The main advantage of free will is the subjective experience of free will, which makes the search for hidden rewards more deliberate and focused, and thus saves time and is more efficient in many situations. Dennett (128) argues that consciously sensing free will is a crucial characteristic of free will. The evolutionary benefits driving the appearance of free will would be the boost in initiative, resulting from the conscious belief that we can deliberately decide our own behavior. This belief is also the basis for moral responsibility and thus helps social coherence. And adding theory of mind would make me understand others' free will and voluntary behavior and further facilitate social processes.

One view on conscious free will assumes that it is generated from scratch and makes us completely free to do what we want. There are only two possibilities to generate this sort of free will. First, brain activity generates the conscious awareness as necessary hallmark of free will. This brain activity necessarily precedes the conscious awareness and thus cannot be directed by free will. In this case, free will would not be free but an illusion of freedom that follows the uncontrolled generation of free will

by the brain. Alternatively, it is not brain activity but something nonmaterial that leads to conscious awareness of free will, and then brain activity carries out what the nonmaterial event has directed it to do and signals free will. Both scenarios seem unresolvable and therefore depressing and ultimately unconstructive. However, the problem may be overcome by taking a step back to consider the neuronal mechanisms by which free will might arise in more realistic situations in which it is seriously constrained by a number of factors. Even in this restricted form, free will would still provide an evolutionary beneficial, conscious control over our actions.

Everyday neurophysiological experiments tell us that many, maybe most, neurons in the brain are spontaneously active without any observable stimuli or actions. Thus brains generate their own activity. We also know that we have many mental experiences that seem to arise spontaneously out of nowhere, even in fully awake and nonpathological states, like thoughts, imagination, and old memories. How could I write this text if I could not sit back (with the proverbial cup of tea) and listen to the ideas that are popping up in my head (most of them rubbish, but occasionally a useful one), and then find the words and put them into grammatically correct order? Neuronal activity in cortex and striatum ramps up over several seconds before self-initiated movements but not before externally instructed actions (299, 311, 337, 380, 397, 491, 493, 526, 528). The key to explaining such brain activity may be the tendency to stochastic activity in many neurons and networks (129, 167, 180, 195). Once initiated by random molecular fluctuations (94, 287, 561, 567), the noise may result at some point in organized activity (168, 200, 297, 330, 465, 564, 611) and influence or settle in attractor states (63, 125, 510) that result in mental events. The hardware of the individual's brain and the individual's experience would constrain the range of attractor states producing mental events. A person who has never done or watched or heard about underwater diving may never experience a spontaneous imagination of the color of corals, nor the free will to go diving to see this. Thus the stochastic brain activity would arise spontaneously, relate to past experiences, and not be initiated from scratch by free will (nor do we need to assume nonmaterial origins).

The spontaneously occurring attractor states arising from initially stochastic neuronal activity would not need to be conscious. Several such elementary unconscious attractor states may develop at the same time, each of them representing a different future option for intention. The possibility of developing attractor states from stochastic activity would not be infinite but depend on brain activity restricted by hardware, current environment, and memory from earlier experience. Thus the attractor states would be restricted to viable options within boundaries. Impulses that never come up cannot be selected. Attractor states related to the intention of exploring the color of corals may only occur if my brain can process color and if I have ever heard about colorful corals. Then a selection mechanism would be set in motion in which viable options emerge with a certain degree of variation within bounds defined by hardware and personal experience. The process may resemble the selection of a reasonable option from a rotating initiator of options. There would be no deliberate, free initiative involved in this selection, as not only the attractor states but also the selection process and its boundaries would be unconsciously mechanical. Only the result of the selection would pop up into consciousness and be interpreted as free will. This conscious free will is not an illusion but the result of unconscious processing. From this point on, it would be possible to decide consciously whether to follow the selected impulse for action or to resist, to select consciously between the emerging mental images and intentions (e.g., separating rubbish from nonrubbish in writing), or to listen further to solutions from those popping up. Thus the unconscious processes produce a scratchpad that pops into consciousness and from which conscious free will can choose freely or veto an intention. This might be what is usually associated with the term of free will and the individual's responsibility for their own intentional actions.

Of course, there may be completely different, currently unknown processes besides stochasticity that may generate brain states underlying free will. The idea would require the discovery of new phenomena in physics (429). However, some of the described steps in the emergence of free will may still hold, in particular the dilemma in the initiation of free will between the brain's hardware and the mind.

2. Constraints on free will

Although free will is associated with being able to make choices (219), I can only choose between the options that are available. We are just not free to choose between anything imaginable, but our options are restricted by the physical state of the world (we cannot fly to Mars), our body's hardware (we cannot fly like birds), and our personal experiences and education (most language scholars do not do partial differential equations). Even among the available options, our choices are influenced by factors not considered as economic value such as emotions driven by external events, social pressure and conventions, altruism, moral rules, and irrational beliefs such as nationalism, chauvinism, prejudices, and superstitions. These external and internal constraints reduce free will in a graded manner, and up to its complete loss when constraints become too serious. Of course, survival is the strongest and most natural constraint on free will. Only within these bounds is free will free.

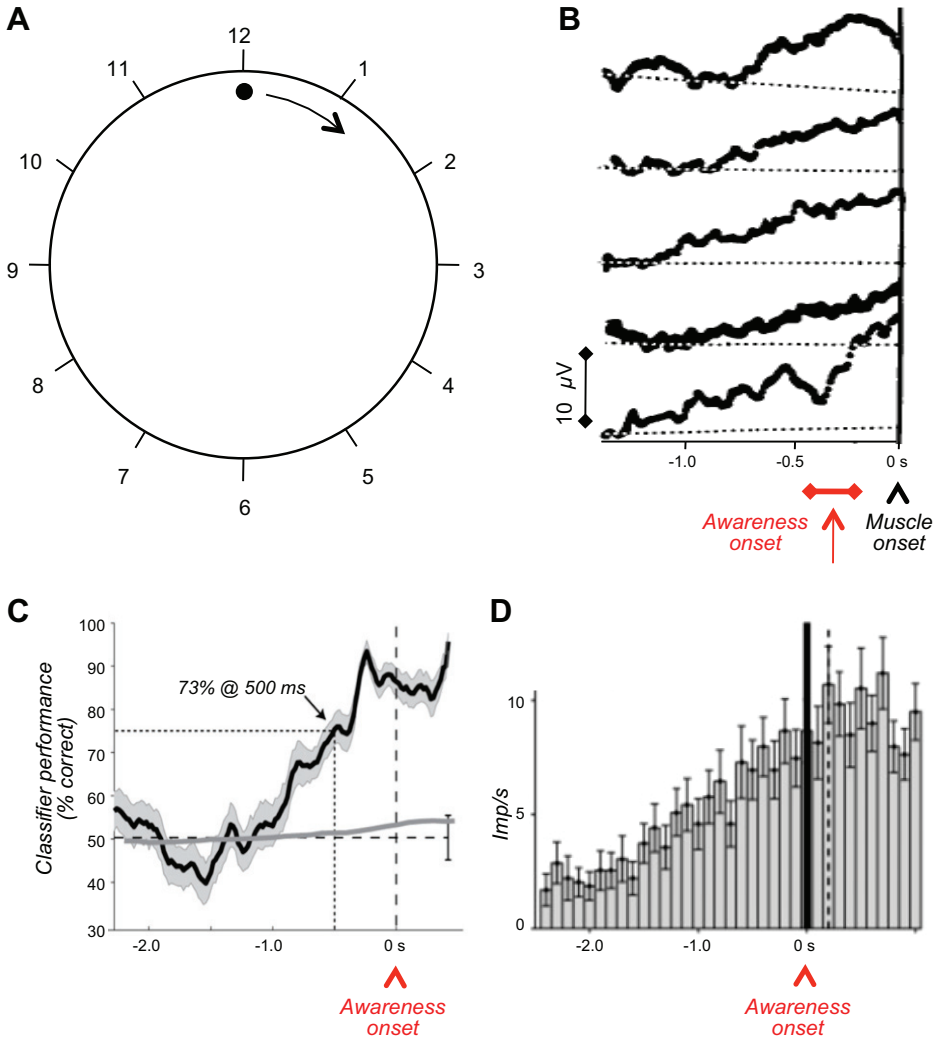


FIGURE 49. Unconscious onset of conscious awareness to act. *A:* design of Libet's classic experiment. A dot moves clockwise around a circle. Subjects remember the dot position at which they became consciously aware of the urge to do a self-initiated finger or wrist movement and report the dot position after being prompted. *B:* electroencephalographic readiness potential recorded at the vertex (top of skull) of one human participant (average of 40 trials). Onset of aware urge to move is indicated in red (bar shows range from eight averages of 40 trials each). Muscle onset denotes onset of electromyographic activity in involved forearm muscle. [From Libet et al. (316), by permission of Oxford University Press.] *C:* detection of time and occurrence of deviation from baseline activity in 37 single and multiple neurons before awareness to act, as determined by a support vector machine classifier. The neurons or multineuron clusters are located in human supplementary motor area (n = 22), anterior cingulate cortex (n = 8), and medial temporal lobe (n = 7). *D:* ramping activity preceding the conscious awareness to act in activity averaged from 59 neurons in human supplementary motor area. [*C* and *D* from Fried et al. (172), with permission from Elsevier.]

Free will can be constrained and focused. A good example is intention in action (529). I choose deliberately a course of action that allows me to form short-lived intentions on the fly while I am moving. I choose to do a walk through town and find a shop window that attracts my intention and makes me buy a pair of boots for the winter. My free will is partly restricted by the decision to take the walk, and the shoe purchase reflects that constraint. My purchase may simply be triggered by the view of the shoes, their price, and their relationship to my purchasing habits and personal preferences. The intention to enter the shop comes instantaneously, almost as a reaction to the sensory input. Very little free will is involved in the final purchasing action after I have initiated the walk.

In a similar situation of very restricted free will, humans indicate the moment at which they consciously perceive the awareness to initiate a simple finger movement **(FIGURE 49*A*)** (316). Indeed, the participants become aware of the urge to move only several hundred milliseconds after the onset of electroencephalographic brain activity for the movement (readiness potential, **FIGURE 49*B***). Sensory and motor confounds are excluded. The only conscious component consists of the possibility to veto the arising urge to move before the actual movement starts. Action-specific preparatory increases of BOLD signals in frontopolar and parieto-cingulate cortex precede the conscious awareness to move by 7–10 s (555). The activity of single neurons in humans preceding self-initiated movements predicts the awareness to move by 500 ms with 70–90% precision **(FIGURE 49*C*)**. These potentials may originate in nonintentional, spontaneous fluctuations reflecting a state of neuronal propensity toward action initiation, true to the meaning of the word "readiness potential" (515). At the neuronal level, such spontaneous fluctuations might occur below action potential threshold but are not conspicuous when recording action potentials during self-initiated movements in the supplementary motor area and associated cortical and subcortical structures in monkeys not expressing awareness to move (299, 311, 337, 380, 397, 491, 493, 526, 528) **(FIGURE 37*C*)** and in the supplementary motor area and anterior cingulate cortex in humans 1–2 s before the awareness to move **(FIGURE 49*D*)** (172). Thus it appears

as if the brain's hardware initiated the movement without the participant's own initiative, suggesting that conscious free will does not exist in these experiments. It is tempting to deny the general existence of free will on the basis of these results. However, the voluntary submission to the experiment may put so many constraints on the liberty of the participant that indeed there is no free will left and the brain is able to initiate the movement without conscious free will, just like the shoe buying example. Thus the experiments employ the role of constraints on free will in an extreme manner and test intention in action rather than true free will. In doing so they may reveal the existence of the elementary process outlined above by which stochastic brain activity settles spontaneously into an attractor state without conscious awareness and only afterwards leads to a conscious mental state. The consciously aware urge to move is then simply a retrograde interpretation of an unconsciously initiated action following the conscious selection to undergo the experiment.

3. Rewards exert constraints on free will

Rewards are necessary for survival. Without rewards we would die of thirst within days and of hunger within weeks, depending on our adipose reserves. With that function, rewards influence and thus constrain behavior and restrict free will in several ways. The passive experience of rewards attracts behavioral reactions away from other objects and focuses behavior onto the available rewards. Entering active decisions between rewards is not very free either because we need rewards at some point for survival. Once we are able to decide, we choose the best option, called utility maximization in economics, to ensure the highest chance for survival and gene propagation. Choices in such restricted situations may often reflect intentions in action, similar to the Libet experiments on movements, and involve serious constraints on free will. The choices are easy to do when the options are limited and have clearly different values, in which case there is little scope for free will, but the choices are more difficult when more, and more similar, options remove some constraints on free will. Habits for acquiring rewards impose particularly strong constraints on free will, and addictions to natural and artificial rewards eliminate free will even more. The initial administration of an addictive substance, like nicotine, is done with free will, often constrained by peer pressure, but following the habit once serious addiction has set in is almost deterministic.

All additional factors affecting reward value add to the constraints on free will. Motivational states like hunger and thirst make rewards more valuable and prioritize our behavior, as easily seen before lunch and dinner. Rewarding contexts make us long for rewards. Who would work during an office party when colleagues drink wine and tell interesting stories? Work required to obtain rewards seriously focuses our actions. Many animals spend most of their day searching for food. Temporal discounting of future rewards hinders our planning and favors myopic actions. Reference-dependent valuation makes us slaves of our environment. Nobody wants to lose what they have achieved, and keeping up with the Joneses is hard work, both of which require good focus and constitute constraints.

Risk exerts huge constrains on our behavior. Risk avoidance makes us shy away from risky situations even if they might be beneficial in the long run. The inability to completely eliminate risk induces anxiety, prejudice, and superstition which further direct behavior in serious ways, up to generating wars without reason. Risk management and the desire for risk reduction are prevalent in all cultures. Many activities of daily life are aimed at reducing risk, from paying car and health insurance to ritual sacrifices. The first reaction of the Minoan people to the devastating Santorini volcano eruption was human sacrifices to appease the angry gods. Risk reduction becomes a part of the culture, influences the whole life history of individuals, and thus takes some free will away.

The constraints exerted by rewards on free will extend also to the learning about rewards. Pavlovian reward predictions arise automatically following repeated exposure without the subject's own doing, which constrains free will substantially. Operant conditioning involves own action and allows goal-directed behavior, which constrains free will a bit less than Pavlovian predictions. Conditioning and higher forms of learning are evolutionary beneficial, as they increase reward acquisition for survival and gene propagation, but in doing so they constrain free will considerably.

Genetic dispositions, education, and long-term reward experience are instrumental for establishing personal reward preferences that influence the whole life of individuals. Different individuals prefer different rewards and choose their profession accordingly (among many other factors). They become mechanics because they appreciate precision, bankers for the money, musicians for the beautiful sounds, doctors because of altruism, and scientists because they value curiosity and knowledge highly. The impact of these preferences accumulates over the whole life span and shapes the personality. All of these influences constrain free will. Even seemingly free changes in personal reward preferences due to physiological factors and personal experience do not much enhance free will.

Punishers constitute the second major class of motivators of behavior besides rewards. Punishers derive from external sources over which we have little control, and they need to be avoided to escape their damaging impact. Thus avoidance is a reaction to events that occurs without one's own doing. There may be different options for escape, and thus limited scope for free will, but overall avoidance involves little free will, allows little deliberation, and produces primarily reactive rather than proactive behavior. The choices

between punishers seem to allow some degree of liberty that is intrinsically low when being forced into the choice. According to the *New York Times* of April 23, 2010, prisoners sentenced to capital punishment in Utah may choose between hanging and a firing squad. "I would like the firing squad, please" reportedly said Ronnie Lee Gardner. Overall, some liberty exists when being able to proactively search for best solutions for efficient avoidance. We can think of known dangers and generate preventive measures, but it is hard to imagine totally unknown dangers and be creative in planning their avoidance. Thus punishers may exert even more constraints on free will than rewards.

4. Neuronal reward processing constrains free will

Electrical and optogenetic stimulation of dopamine neurons are very efficient for eliciting immediate behavioral reactions and behavioral learning **(FIGURE 19, *A–D*)** (2, 103, 155, 562, 605, 641). The strength of dopamine stimulation affects choices, reaction times, and reward acquisition in a graded manner. The animals are simply driven towards goals associated with the dopamine activation, both during the stimulation (immediate effect) and in the next test sessions (reinforcement learning). Thus, in focusing individuals on rewards at the expense of other pursuits, dopamine stimulation takes parts of free will away from the individual's behavior.

Dopamine stimulations are not experimental artifacts but mimic the natural phasic dopamine responses to reward prediction errors. The dopamine responses are likely to have similar behavioral effects as the stimulation. Even though stimulation parameters differ somewhat from the natural responses, the important feature is the brief, phasic dopamine activation, whereas slow, tonic dopamine stimulation is ineffective (e.g., for place preference learning; Ref. 605). The dopamine effects likely involve synaptic plasticity and focusing in striatum and frontal cortex (70, 368, 424, 540, 594). Indeed, optogenetic stimulation of striatal neurons postsynaptic to dopamine neurons induces differential approach and avoidance learning and behavioral bias depending on the dopamine receptor subtype the stimulated neurons carry **(FIGURES 19*F* AND 46*C*)** (293, 576). Thus the combined evidence from dopamine stimulation and dopamine reward responses indicates that eliciting a strong reward signal in dopamine neurons directs the organism towards reward and thus constrains its free will. This is the mechanism through which primary (unpredicted) rewards and (Pavlovian) reward predictors influence the behavior and constrain free will.

A similar argument may hold for neuronal risk processing. Dopamine and orbitofrontal reward neurons code risk separately from value **(FIGURE 31*C*)** (161, 162, 391, 392). Furthermore, neuronal risk signals, and the influence of risk on neuronal value signals, vary with individual risk attitudes (301). If these risk signals could affect risk components in behavior analogous to value signals, they would provide mechanistic explanations for the constraining influence of risk on free will. To demonstrate such an effect, one would need to demonstrate that selective stimulation of risk neurons, such as those found in orbitofrontal cortex, affects risk attitudes during choices.

The strength of the dopamine response seems to convey the influence of rewards on behavior. Rewards with stronger effects on dopamine neurons are likely to have more impact on learning and choices. Thus the propensity of individual rewards to activate dopamine neurons would determine the influence of reward on behavioral choices. This function may hold also for other reward centers in the brain, although their less compact organization may not be conducive to such dramatic stimulation effects. However, it is unlikely that a response of dopamine neurons or other reward neurons to an automatically conditioned, reward-predicting stimulus induces approach behavior entirely without the subject's own doing. There are other brain mechanisms that would limit the automaticity of such effects, but any neuronal reward signal is nevertheless likely to have a biasing effect towards rewarded stimuli and actions. The neuronal responsiveness to different rewards is likely to vary between individuals. Interindividual differences in reward processing may affect daily preferences but also determine long-term behavior, including professional choices and other important decisions in life. Thus the activity of reward neurons shapes behavior, constrains voluntary decisions, and thus restricts free will.

ACKNOWLEDGMENTS

I thank Anthony Dickinson for 20 years of discussion and collaboration on animal learning theory, Peter Bossaerts and Christopher Harris for 10 years of discussion and collaboration on experimental economics, and Peter Bossaerts, Anthony Dickinson, Fabian Grabenhorst, Armin Lak, Johannes Schultz, William R. Stauffer and two anonymous referees for comments. I am indebted to Colin F. Camerer, David M. Grether, John O. Ledyard, Charles R. Plott, and Antonio Rangel for discussing experimental economics at the Division of Humanities and Social Sciences, California Institute of Technology, Pasadena.

Address for reprint requests and other correspondence: W. Schultz, Dept. of Physiology, Development and Neuroscience, Univ. of Cambridge, Cambridge CB2 3DY, UK (e-mail: ws234@cam.ac.uk).

GRANTS

This work is supported by Wellcome Trust Principal Research Fellowship, Programme and Project Grants 058365, 093270, and 095495; European Research Council Advanced Grant 293549; and National Institutes of Health Caltech Conte Center Grant P50MH094258. Previous sup-

port came from the Human Frontiers Science Program, the Wellcome-MRC-funded Behavioural and Clinical Neuroscience Institute Cambridge, and the Swiss National Science Foundation.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author.

REFERENCES

1. Abe H, Lee D. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron* 70: 731–741, 2011.
2. Adamantidis AR, Tsai HC, Boutrel B, Zhang F, Stuber GD, Budygin EA, Touriño C, Bonci A, Deisseroth K, de Lecea L. Optogenetic interrogation of dopaminergic modulation of the multiple phases of reward-seeking behavior. *J Neurosci* 31: 10829–10835, 2011.
3. Adrian ED, Zotterman Y. The impulses produced by sensory nerve endings. Part 3. Impulses set up by Touch and Pressure. *J Physiol* 61: 465–483, 1926.
4. Ainslie GW. Impulse control in pigeons. *J Exp Anal Behav* 21: 485–489, 1974.
5. Ainslie GW. Specious rewards: a behavioral theory of impulsiveness and impulse control. *Psych Bull* 82: 463–496, 1975.
6. Alexander GE, Crutcher MD. Neural representations of the target (goal) of visually guided arm movements in three motor areas of the monkey. *J Neurophysiol* 64: 164–178, 1990.
7. Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9: 357–381, 1986.
8. Amador N, Schlag-Rey M, Schlag J. Reward-predicting and reward-detecting neuronal activity in the primate supplementary eye field. *J Neurophysiol* 84: 2166–2170, 2000.
9. Ambroggi F, Ishikawa A, Fields HL, Nicola SM. Basolateral amygdala neurons facilitate reward-seeking behavior by exciting nucleus accumbens neurons. *Neuron* 59: 648–661, 2008.
10. Amemori KI, Graybiel AM. Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. *Nat Neurosci* 15: 776–785, 2012.
11. Amiez C, Joseph JP, Procyk E. Anterior cingulate error-related activity is modulated by predicted reward. *Eur J Neurosci* 21: 3447–3452, 2005.
12. Amiez C, Joseph JP, Procyk E. Reward encoding in the monkey anterior cingulate cortex. *Cereb Cortex* 16: 1040–1055, 2006.
13. Aosaki T, Graybiel AM, Kimura M. Effect of the nigrostriatal dopamine system on acquired neural responses in the striatum of behaving monkeys. *Science* 265: 412–415, 1994.
14. Aosaki T, Tsubokawa H, Ishida A, Watanabe K, Graybiel AM, Kimura M. Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *J Neurosci* 14: 3969–3984, 1994.
15. Apicella P, Deffains M, Ravel S, Legallet E. Tonically active neurons in the striatum differentiate between delivery and omission of expected reward in a probabilistic task context. *Eur J Neurosci* 30: 515–526, 2009.
16. Apicella P, Ljungberg T, Scarnati E, Schultz W. Responses to reward in monkey dorsal and ventral striatum. *Exp Brain Res* 85: 491–500, 1991.
17. Apicella P, Ravel S, Deffains M, Legallet E. The role of striatal tonically active neurons in reward prediction error signaling during instrumental task performance. *J Neurosci* 31: 1507–1515, 2011.
18. Apicella P, Scarnati E, Ljungberg T, Schultz W. Neuronal activity in monkey striatum related to the expectation of predictable environmental events. *J Neurophysiol* 68: 945–960, 1992.
19. Argilli E, Sibley DR, Malenka RC, England PM, Bonci A. Mechanism and time course of cocaine-induced long-term potentiation in the ventral tegmental area. *J Neurosci* 28: 9092–9100, 2008.
20. Arsenault JT, Rima S, Stemmann H, Vanduffel W. Role of the primate ventral tegmental area in reinforcement and motivation. *Curr Biol* 24: 1347–1353, 2014.
21. Asaad WF, Eskandar EN. Encoding of both positive and negative reward prediction errors by neurons of the primate lateral prefrontal cortex and caudate nucleus. *J Neurosci* 31: 17772–17787, 2011.
22. Aston-Jones G, Rajkowski J, Kubiak P, Alexinsky T. Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. *J Neurosci* 14: 4467–4480, 1994.
23. Azzi JCB, Sirigu A, Duhamel JR. Modulation of value representation by social context in the primate orbitofrontal cortex. *Proc Natl Acad Sci USA* 109: 2126–2131, 2012.
24. Badman MK, Flier JS. The gut and energy balance: visceral allies in the obesity wars. *Science* 307: 1909–1914, 2005.
25. Báez-Mendoza R, Harris C, Schultz W. Activity of striatal neurons reflects social action and own reward. *Proc Natl Acad Sci USA* 110: 16634–16639, 2013.
26. Balleine B, Dickinsion A. Goal-directed instrumental action: contingency, and incentive learning and their cortical substrates. *Neuropharmacology* 37: 407–419, 1998.
27. Bangasser DA, Waxler DE, Santollo J, Shors TJ. Trace conditioning and the hippocampus: the importance of contiguity. *J Neurosci* 26: 8702–8706, 2006.
28. Bao S, Chan VT, Merzenich MM. Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature* 412: 79–83, 2001.
29. Barnes TD, Kubota Y, Hu D, Jin DZ, Graybiel AM. Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* 437: 1158–1161, 2005.
30. Barraclough D, Conroy ML, Lee DJ. Prefrontal cortex and decision making in a mixed-strategy game. *Nat Neurosci* 7: 405–410, 2004.
31. Barto AG, Singh S, Chentanez N. Intrinsically motivated learning of hierarchical collections of skills. In: *Advances in Neural Information Processing Systems 17: Proceedings of the 2004 Conference*. Cambridge, MA: MIT Press, 2005.
32. Barto AG, Sutton RS, Anderson CW. Neuronlike adaptive elements that can solve difficult learning problems. *IEEE Trans Syst Man Cybernet* 13: 834–846, 1983.
33. Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47: 129–141, 2005.
34. Bayer HM, Lau B, Glimcher PW. Statistics of dopamine neuron spike trains in the awake primate. *J Neurophysiol* 98: 1428–1439, 2007.
35. Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 50: 7–15, 1994.
36. Belova MA, Paton JJ, Morrison SE, Salzman CD. Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron* 55, 970–984, 2007.
37. Belova MA, Paton JJ, Salzman CD. Moment-to-moment tracking of state value in the amygdala. *J Neurosci* 28: 10023–10030, 2008.
38. Bennur S, Gold JI. Distinct representations of a perceptual decision and the associated oculomotor plan in the monkey lateral intraparietal area. *J Neurosci* 31: 913–921, 2011.
39. Bentham J. *An Introduction to the Principle of Morals and Legislations*, *1789* . Reprinted Oxford, UK: Blackwell, 1948.
40. Bermudez MA, Göbel C, Schultz W. Sensitivity to temporal reward structure in amygdala neurons. *Curr Biol* 22: 1839–1844, 2012.
41. Bermudez MA, Schultz W. Responses of amygdala neurons to positive reward predicting stimuli depend on background reward (contingency) rather than stimulus-reward pairing (contiguity). *J Neurophysiol* 103: 1158–1170, 2010.
42. Bermudez MA, Schultz W. Reward magnitude coding in primate amygdala neurons. *J Neurophysiol* 104: 3424–2432, 2010.

43. Bernoulli D. Specimen theoriae novae de mensura sortis. *Comentarii Academiae Scientiarum Imperialis Petropolitanae (Papers Imp Acad Sci St Petersburg)* 5: 175–192, 1738 (translated as Exposition of a new theory on the measurement of risk. *Econometrica* 22: 23–36, 1954).

44. Berridge KC. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology* 191: 391–431, 2007.

45. Berridge KC, Robinson TE. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res Rev* 28: 309–369, 1998.

46. Berthoud HR, Morrison C. The brain, appetite, obesity. *Annu Rev Psychol* 59: 55–92, 2008.

47. Beylin AV, Gandhi CC, Wood GE, Talk AC, Matzel LD, Shors TJ. The role of the hippocampus in trace conditioning: temporal discontinuity or task difficulty? *Neurobiol Learn Mem* 76: 447–461, 2001.

48. Bi GQ, Poo MM. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci* 18: 10464–10472, 1998.

49. Bichot NP, Schall JD. Effects of similarity and history on neural mechanisms of visual selection. *Nat Neurosci* 2: 549–554, 1999.

50. Binmore K, Shaked A. Experimental economics: where next? *J Econ Behav Organiz* 73: 87–100, 2010.

51. Black RW. Shifts in magnitude of reward and contrast effects in instrumental and selective learning: a reinterpretation. *Psychol Rev* 75: 114–126, 1968.

52. Blanchard TC, Wolfe LS, Vlaev I, Winston JS, Hayden BY. Biases in preferences for sequences of outcomes in monkeys. *Cognition* 130: 289–299, 2014.

53. Blatter K, Schultz W. Rewarding properties of visual stimuli. *Exp Brain Res* 168: 541–546, 2006.

54. Blythe SN, Wokosin D, Atherton JF, Bevan MD. Cellular mechanisms underlying burst firing in substantia nigra dopamine neurons. *J Neurosci* 29: 15531–15541, 2009.

55. Bogacz R. Optimal decision-making theories: linking neurobiology with behaviour. *Trends Cog Sci* 11: 118–125, 2007.

56. Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev* 113: 700–765, 2006.

57. Bogacz R, Gurney K. The basal ganglia and cortex implement optimal decision making between alternative actions. *Neur Comput* 19: 442–477, 2007.

58. Bouret S, Richmond BJ. Ventromedial and orbital prefrontal neurons differentially encode internally and externally driven motivational values in monkeys. *J Neurosci* 30: 8591–8601, 2010.

59. Bouton ME. Context and behavioural processes in extinction. *Learning Memory* 11: 485–494, 2004.

60. Bowman EM, Aigner TG, Richmond BJ. Neural signals in the monkey ventral striatum related to motivation for juice and cocaine rewards. *J Neurophysiol* 75: 1061–1073, 1996.

61. Bredfeldt CE, Ringach DL. Dynamics of spatial frequency tuning in macaque V1. *J Neurosci* 22: 1976–1984, 2002.

62. Breton YA, James C, Marcus JC, Shizgal P. *Rattus psychologicus*: construction of preferences by self-stimulating rats. *Behav Brain Res* 202: 77–91, 2009.

63. Brzezniak Z, Capinski M, Flandoli F. Pathwise global attractors for stationary random dynamical systems. *Probab Theory Relat Fields* 95: 87–102, 1993.

64. Brischoux F, Chakraborty S, Brierley DI, Ungless MA. Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. *Proc Natl Acad Sci USA* 106: 4894–4899, 2009.

65. Brody CD, Hernandez A, Zainos A, Romo R. Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb Cortex* 13: 1196–1207, 2003.

66. Bromberg-Martin ES, Hikosaka O. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron* 63: 119–126, 2009.

67. Bromberg-Martin ES, Hikosaka O. Lateral habenula neurons signal errors in the prediction of reward information. *Nat Neurosci* 14: 1209–1216, 2011.

68. Bromberg-Martin ES, Matsumoto M, Hon S, Hikosaka O. A pallidus-habenula-dopamine pathway signals inferred stimulus values. *J Neurophysiol* 104: 1068–1076, 2010.

69. Brosnan SF, de Waal FBM. Monkeys reject unequal pay. *Nature* 425: 297–299, 2003.

70. Brown JR, Arbuthnott GW. The electrophysiology of dopamine (D_2) receptors: a study of the actions of dopamine on corticostriatal transmission. *Neuroscience* 10: 349–355, 1983.

71. Brown JW, Bullock D, Grossberg S. How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *J Neurosci* 19: 10502–10511, 1999.

72. Brown JW, Bullock D, Grossberg S. How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks* 17: 471–510, 2004.

73. Brown MTC, Henny P, Bolam JP, Magill PJ. Activity of neurochemically heterogeneous dopaminergic neurons in the substantia nigra during spontaneous and driven changes in brain state. *J Neurosci* 29: 2915–2925, 2009.

74. Bruce CJ, Goldberg ME. Primate frontal eye fields. I. Single neurons discharging before saccades. *J Neurophysiol* 53: 603–635, 1985.

75. Budygin EA, Park J, Bass CE, Grinevich VP, Bonin KD, Wightman RM. Aversive stimulus differentially triggers subsecond dopamine release in reward regions. *Neuroscience* 201: 331–337, 2012.

76. Bunney BS, Grace AA. Acute and chonic haloperidol treatment: comparison of effects on nigral dopaminergic cell activity. *Life Sci* 23: 1715–1728, 1978.

77. Burkart JM, Fehr E, Efferson C, van Schaik CP. Other-regarding preferences in a non-human primate: common marmosets provision food altruistically. *Proc Natl Acad Sci USA* 104: 19762–19766, 2007.

78. Caggiano V, Fogassi L, Rizzolatti G, Casile A, Giese MA, Thier P. Mirror neurons encode the subjective value of an observed action. *Proc Natl Acad Sci USA* 109: 11848–11853, 2012.

79. Cai X, Kim S, Lee D. Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. *Neuron* 69: 170–182, 2011.

80. Cai X, Padoa-Schioppa C. Neuronal encoding of subjective value in dorsal and ventral anterior cingulate cortex. *J Neurosci* 32: 3791–3808, 2012.

81. Cai X, Padoa-Schioppa C. Contributions of orbitofrontal and lateral prefrontal cortices to economic choice and the good-to-action transformation. *Neuron* 81: 1140–1151, 2014.

82. Calabresi P, Gubellini P, Centonze D, Picconi B, Bernardi G, Chergui K, Svenningsson P, Fienberg AA, Greengard P. Dopamine and cAMP-regulated phosphoprotein 32 kDa controls both striatal long-term depression and long-term potentiation, opposing forms of synaptic plasticity. *J Neurosci* 20: 8443–8451, 2000.

83. Camerer CF. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton Univ. Press, 2003.

84. Camille N, Tsuchida A, Fellows LK. Double dissociation of stimulus-value and action-value learning in humans with orbitofrontal or anterior cingulate cortex damage. *J Neurosci* 31: 15048–15052, 2001.

85. Caplin A, Dean M, Martin D. Search and satisficing. *Am Econ Rev* 101: 2899–2922, 2011.

86. Caraco T, Blankenhorn WU, Gregory GM, Newman JA, Recer GM, Zwicker SM. Risk-sensitivity: ambient temperature affects foraging choice. *Anim Behav* 39: 338–345, 1990.

87. Caraco T, Martindale S, Whitham TS. An empirical demonstration of risk-sensitive foraging preferences. *Anim Behav* 28: 820–830, 1980.

88. Cardinal RN, Howes NJ. Effects of lesions of the nucleus accumbens core on choice between small certain and large uncertain rewards in rats. *BMC Neurosci* 6: 37, 2005.

89. Chang SWC, Gariépy JF, Platt ML. Neuronal reference frames for social decisions in primate frontal cortex. *Nat Neurosci* 16: 243–250, 2013.

90. Chang SWC, Winecoff AA, Platt ML. Vicarious reinforcement in rhesus macaques (*Macaca mulatta*). *Front Neurosci* 5: 27, 2011.

91. Chen MK, Lakshminarayanan V, Santos LR. How basic are behavioral biases? Evidence from capuchin trading behavior. *J Pol Econ* 114: 517–537, 2006.

92. Chesselet MF. Presynaptic regulation of neurotransmitter release in the brain: facts and hypothesis. *Neuroscience* 12: 347–375, 1984.

93. Chevalier G, Vacher S, Deniau JM, Desban M. Disinhibition as a basic process in the expression of striatal functions. I. The striato-nigral influence on tecto-spinal/tecto-diencephalic neurons. *Brain Res* 334: 215–226, 1985.

94. Chow CC, White JA. Spontaneous action potentials due to channel fluctuations. *Biophys J* 71: 3013–3023, 2000.

95. Christoph GR, Leonzio RJ, Wilcox KS. Stimulation of the lateral habenula inhibits dopamine-containing neurons in the substantia nigra and ventral tegmental area of the rat. *J Neurosci* 6: 613–619, 1986.

96. Churchland AK, Kiani R, Chaudhuri R, Wang XJ, Pouget A, Shadlen MN. Variance as a signature of neural computations during decision making. *Neuron* 69: 818–831, 2011.

97. Churchland AK, Kiani R, Shadlen MN. Decision-making with multiple alternatives. *Nat Neurosci* 11: 693–702, 2008.

98. Cisek P, Kalaska JF. Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron* 45: 801–814, 2005.

99. Cisek P, Kalaska JF. Neural mechanisms for interacting with a world full of action choices. *Annu Rev Neurosci* 33: 269–298, 2010.

100. Clarke HF, Dalley JW, Crofts HS, Robbins TW, Roberts AC. Cognitive inflexibility after prefrontal serotonin depletion. *Science* 304: 878–880, 2004.

101. Coe B, Tomihara K, Matsuzawa M, Hikosaka O. Visual and anticipatory bias in three cortical eye fields of the monkey during an adaptive decision-making task. *J Neurosci* 22: 5081–5090, 2002.

102. Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482: 85–88, 2012.

103. Corbett D, Wise RA. Intracranial self-stimulation in relation to the ascending dopaminergic systems of the midbrain: a moveable microelectrode study. *Brain Res* 185: 1–15, 1980.

104. Crespi LP. Quantitative variation in incentive and performance in the white rat. *Am J Psychol* 40: 467–517, 1942.

105. Critchley HG, Rolls ET. Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *J Neurophysiol* 75: 1673–1686, 1996.

106. Cromwell HC, Hassani OK, Schultz W. Relative reward processing in primate striatum. *Exp Brain Res* 162: 520–525, 2005.

107. Cromwell HC, Schultz W. Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *J Neurophysiol* 89: 2823–2838, 2003.

108. Croxson PL, Walton ME, O'Reilly JX, Behrens TEJ, Rushworth MFS. Effort-based cost-benefit valuation and the human brain. *J Neurosci* 29: 4531–4541, 2009.

109. Cui H, Andersen RA. Posterior parietal cortex encodes autonomously selected motor plans. *Neuron* 56: 552–559, 2007.

110. d'Acremont M, Fornari E, Bossaerts P. Activity in inferior parietal and medial prefrontal cortex signals the accumulation of evidence in a probability learning task. *PLoS Comput Biol* 9: e1002895, 2013.

111. d'Ardenne K, McClure SM, Nystrom LE, Cohen JD. BOLD Responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* 319: 1264–1267, 2008.

112. Darwin C. *On the Origin of Species by Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray, 1859.

113. Daw ND, Courville AC, Touretsky DS. Representation and timing in theories of the dopamine system. *Neur Comput* 18: 1637–1677, 2006.

114. Dawkins R. *The Selfish Gene*. London: Oxford Univ. Press, 1976.

115. Day JJ, Jones JL, Carelli RM. Nucleus accumbens neurons encode predicted and ongoing reward costs in rats. *Eur J Neurosci* 33: 308–321, 2011.

116. Day JJ, Jones JL, Wightman RM, Carelli RM. Phasic nucleus accumbens dopamine release encodes effort- and delay-related costs. *Biol Psychiat* 68: 306–309, 2010.

117. Day JJ, Roitman MF, Wightman RM, Carelli RM. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat Neurosci* 10: 1020–1028, 2007.

118. De Araujo IE, Oliveira-Maia AJ, Sotnikova TD, Gainetdinov RR, Caron MC, Nicolelis MAL, Simon SA. Food reward in the absence of taste receptor signaling. *Neuron* 57: 930–941, 2008.

119. De Araujo IE, Rolls ET, Velazco MI, Margot C, Cayeux I. Cognitive modulation of olfactory processing. *Neuron* 46: 671–679, 2005.

120. De Lafuente O, Romo R. Neuronal correlates of subjective sensory experience. *Nat Neurosci* 8: 1698–1703, 2005.

121. De Lafuente O, Romo R. Neural correlate of subjective sensory experience gradually builds up across cortical areas. *Proc Natl Acad Sci USA* 103: 14266–14271, 2006.

122. De Lafuente O, Romo R. Dopamine neurons code subjective sensory experience and uncertainty of perceptual decisions. *Proc Natl Acad Sci USA* 49: 19767–19771, 2011.

123. De Waal FB, Davis JM. Capuchin cognitive ecology: cooperation based on projected returns. *Neuropsychologia* 41: 221–228, 2003.

124. Deaner RO, Khera AV, Platt ML. Monkeys pay per view: adaptive valuation of social images by rhesus monkeys. *Curr Biol* 15: 543–548, 2005.

125. Deco G, Rolls ET, Romo R. Stochastic dynamics as a principle of brain function. *Prog Neurobiol* 88: 1–16, 2009.

126. Delamater AR. Outcome selective effects of intertrial reinforcement in a Pavlovian appetitive conditioning paradigm with rats. *Anim Learn Behav* 23: 31–39, 1995.

127. Deniau JM, Chevalier G. Disinhibition as a basic process in the expression of striatal function. II. The striato-nigral influence on thalamocortical cells of the ventromedial thalamic nucleus. *Brain Res* 334: 227–233, 1985.

128. Dennett DC. *Elbow Room. The Varieties of Free Will Worth Wanting*. Boston: MIT Press, 1984.

129. Destexhe A, Contreras D. Neuronal computations with stochastic network states. *Science* 314, 85–88, 2006.

130. Di Ciano P, Cardinal RN, Cowell RA, Little SJ, Everitt B. Differential involvement of NMDA, AMPA/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of Pavlovian approach behavior. *J Neurosci* 21: 9471–77, 2001.

131. Di Loreto S, Florio T, Scarnati E. Evidence that a non-NMDA receptor is involved in the excitatory pathway from the pedunculopontine region to nigrostriatal dopaminergic neurons. *Exp Brain Res* 89: 79–86, 1992.

132. Dickinson A. *Contemporary Animal Learning Theory*. Cambridge, UK: Cambridge Univ. Press, 1980, p. 43.

133. Dickinson A, Balleine B. Motivational control of goal-directed action. *Anim Learn Behav* 22: 1–18, 1994.

134. Ding L, Gold JI. Caudate encodes multiple computations for perceptual decisions. *J Neurosci* 30: 15747–15759, 2010.

135. Ding L, Hikosaka O. Comparison of reward modulation in the frontal eye field and caudate of the macaque. *J Neurosci* 26: 6695–6703, 2006.

136. Dormont JF, Conde H, Farin D. The role of the pedunculopontine tegmental nucleus in relation to conditioned motor performance in the cat. I. Context-dependent and reinforcement-related single unit activity. *Exp Brain Res* 121: 401–10, 1998.

137. Dorris MC, Glimcher PW. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* 44: 365–378, 2004.

138. Doucet G, Descarries L, Garcia S. Quantification of the dopamine innervation in adult rat neostriatum. *Neuroscience* 19: 427–445, 1986.

139. Doya K, Samejima K, Katagiri K, Kawato M. Multiple model-based reinforcement learning. *Neural Comput* 14: 1347–1369, 2002.

140. Drevets WC, Gautier C, Price JC, Kupfer DJ, Kinahan PE, Grace AA, Price JL, Mathis CA. Amphetamine-induced dopamine release in human ventral striatum correlates with euphoria. *Biol Psychiatry* 49: 81–96, 2001.

141. Eblen F, Graybiel AM. Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. *J Neurosci* 15: 5999–6013, 1995.

142. Edgeworth F. *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. New York: Augustus M. Kelly, 1881.

143. Ekman P. An argument for basic emotions. *Cogn Emot* 6: 169–200, 1992.

144. Enomoto K, Matsumoto N, Nakai S, Satoh T, Sato TK, Ueda Y, Inokawa H, Haruno M, Kimura M:. Dopamine neurons learn to encode the long-term value of multiple future rewards. *Proc Natl Acad Sci USA* 108: 15462–15467, 2011.

145. Estes WK. Discriminative conditioning. I: a discriminative property of conditioned anticipation. *J Exp Psychol* 32: 150–155, 1943.

146. Everitt BJ, Stacey P. Studies of instrumental behavior with sexual reinforcement in male rats (*Rattus norvegicus*). II. Effects of preoptic area lesions, castration, and testosterone. *J Com Psychol* 101: 407–419, 1987.

147. Everling S, Munoz DP. Neuronal correlates for preparatory set associated with prosaccades and anti-saccades in the primate frontal eye field. *J Neurosci* 20: 387–400, 2000.

148. Fairhall AL, Lewen GD, Bialek W, de Ruyter van Steveninck RR. Efficiency and ambiguity in an adaptive neural code. *Nature* 412: 787, 2001.

149. Fehr E, Camerer CF. Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn Neurosci* 11: 419–427, 2007.

150. Fehr E, Schmidt KM. A theory of fairness, competition, and cooperation. *Q J Econ* 114: 817–688, 1999.

151. Fehr E, Schmidt KM. On inequity aversion: a reply to Binmore and Shaked. *J Econ Behav Organiz* 73: 101–108, 2010.

152. Fehr-Duda H, Bruhin A, Epper T, Schubert R. Rationality on the rise: Why relative risk aversion increases with stake size. *J Risk Uncertain* 40: 147–180, 2010.

153. Feierstein CE, Quirk MC, Uchida N, Sosulski DL, Mainen ZF. Representation of spatial goals in rat orbitofrontal cortex. *Neuron* 51: 495–507, 2006.

154. Fiala J, Grossberg S, Bullock D. Metabotropic glutamate receptor activation in cerebellar purkinje cells as substrate for adaptive timing of the classically conditioned eye-blink response. *J Neurosci* 16: 3760–3774, 1996.

155. Fibiger HC, LePiane FG, Jakubovic A, Phillips AG. The role of dopamine in intracranial self-stimulation of the ventral tegmental area. *J Neurosci* 7: 3888–3896, 1987.

156. Fiorillo CD. Transient activation of midbrain dopamine neurons by reward risk. *Neuroscience* 197: 162–171, 2011.

157. Fiorillo CD. Two dimensions of value: dopamine neurons represent reward but not aversiveness. *Science* 341: 546–549, 2013.

158. Fiorillo CD, Newsome WT, Schultz W. The temporal precision of reward prediction in dopamine neurons. *Nat Neurosci* 11: 966–973, 2008.

159. Fiorillo CD, Song MR, Yun SR. Diversity and homogeneity in responses of midbrain dopamine neurons. *J Neurosci* 33: 4693–4709, 2013.

160. Fiorillo CD, Song MR, Yun SR. Multiphasic temporal dynamics in responses of midbrain dopamine neurons to appetitive and aversive stimuli. *J Neurosci* 33: 4710–4725, 2013.

161. Fiorillo CD, Tobler PN, Schultz W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299: 1898–1902, 2003.

162. Fiorillo CD, Tobler PN, Schultz W. Evidence that the delay-period activity of dopamine neurons corresponds to reward uncertainty rather than backpropagating TD errors. *Behav Brain Funct* 1: 7, 2005.

163. Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, Akers CA, Clinton SM, Phillips PE, Akil H. A selective role for dopamine in stimulus-reward learning. *Nature* 469: 53–57, 2011.

164. Flaherty AW, Graybiel A. Output architecture of the primate putamen. *J Neurosci* 13: 3222–3237, 1993.

165. Fliessbach K, Weber B, Trautner P, Dohmen T, Sunde U, Elger CE, Falk A. Social comparison affects reward-related brain activity in the human ventral striatum. *Science* 318: 1305–1308, 2007.

166. Floresco SB, West AR, Ash B, Moore H, Grace AA. Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nat Neurosci* 6: 968–973, 2003.

167. Fourcaud-Trocme N, Brunel N. Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neur Comp* 14: 2057–2110, 2002.

168. Fox RF, Lu Y. Emergent collective behavior in large numbers of globally coupled independently stochastic ion channels. *Phys Rev E* 49: 3421–3425, 1994.

169. Frank MJ, Seeberger LC, O'Reilly RC. By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science* 306: 1940–1943, 2004.

170. Frémaux N, Sprekeler H, Gerstner W. Functional requirements for reward-modulated spike-timing-dependent plasticity. *J Neurosci* 30: 13326–13337, 2010.

171. Freund TF, Powell JF, Smith AD. Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spines. *Neuroscience* 13: 1189–1215, 1984.

172. Fried I, Mukamel R, Kreiman G. Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* 69: 548–562, 2011.

173. Friedman M, Savage LJ. The utility analysis of choices involving risk. *J Polit Econ* 56: 279–304, 1948.

174. Gaffan D, Murray EA, Fabre-Thorpe M. Interaction of the amygdala with the frontal lobe in reward memory. *Eur J Neurosci* 5: 968–975, 1993.

175. Gallistel CR, Gibbon J. Time, rate and conditioning. *Psych Rev* 107: 289–344, 2000.

176. Gao WY, Lee TH, King GR, Ellinwood EH. Alterations in baseline activity and quinpirole sensitivity in putative dopamine neurons in the substantia nigra and ventral tegmental area after withdrawal from cocaine pretreatment. *Neuropsychopharmacology* 18: 222–232, 1998.

177. Gauthier J, Parent M, Lévesque M, Parent A. The axonal arborization of single nigrostriatal neurons in rats. *Brain Res* 834: 228–232, 1999.

178. Gdowski MJ, Miller LE, Parrish T, Nenonene EK, Houk JC. Context dependency in the globus pallidus internal segment during targeted arm movements. *J Neurophysiol* 85: 998–1004, 2001.

179. Genovesio A, Tsujimoto S, Navarra G, Falcone R, Wise SP. Autonomous encoding of irrelevant goals and outcomes by prefrontal cortex neurons. *J Neurosci* 34: 1970–1978, 2014.

180. Gerstein G, Mandelbrot B. Random walk models for the spike activity of a single neuron. *Biophys J* 4: 41–68, 1964.

181. Gietzen DW, Hao S, Anthony TG. Mechanisms of food intake repression in indispensable amino acid deficiency. *Annu Rev Nutr* 27: 63–78, 2007.

182. Gilbert PFC, Thach WT. Purkinje cell activity during motor learning. *Brain Res* 128: 309–328, 1977.

183. Glimcher PW. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proc Natl Acad Sci USA* 108: 15647–15654, 2011.

184. Gnadt JW, Andersen RA. Memory related motor planning activity in posterior parietal cortex of macaque. *Exp Brain Res* 70: 216–220, 1988.

185. Gold JI, Shadlen MN. The neural basis of decision making. *Annu Rev Neurosci* 30: 535–574, 2007.

186. Goldman-Rakic PS, Leranth C, Williams MS, Mons N, Geffard M. Dopamine synaptic complex with pyramidal neurons in primate cerebral cortex. *Proc Natl Acad Sci USA* 86: 9015–9019, 1989.

187. Gonzalez R, Wu G. On the shape of the probability weighting function. *Cogn Psychol* 38: 129–166, 1999.

188. Grabenhorst F, Hernadi I, Schultz W. Prediction of economic choice by primate amygdala neurons. *Proc Natl Acad Sci USA* 109: 18950–18955, 2012.

189. Grabenhorst F, Rolls ET, Bilderbeck A. How cognition modulates affective responses to taste and flavor: top-down influences on the orbitofrontal and pregenual cingulate cortices. *Cereb Cortex* 18: 1549–1559, 2008.

190. Grabenhorst F, Rolls ET, Margot C, da Silva MAAP, Velazco MI. How pleasant and unpleasant stimuli combine in different brain regions: odor mixtures. *J Neurosci* 27: 13532–13540, 2007.

191. Grether WF. Pseudo-conditioning without paired stimulation encountered in attempted backward conditioning. *Comp Psychol* 25: 91–96, 1938.

192. Groves PM, Garcia-Munoz M, Linder JC, Manley MS, Martone ME, Young SJ. Elements of the intrinsic organization and information processing in the neostriatum. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, Beiser DG. Cambridge, MA: MIT Press, 1995, p. 51–96.

193. Guarraci FA, Kapp BS. An electrophysiological characterization of ventral tegmental area dopaminergic neurons during differential pavlovian fear conditioning in the awake rabbit. *Behav Brain Res* 99: 169–179, 1999.

194. Gurden H, Takita M, Jay TM. Essential role of D1 but not D2 receptors in the NMDA receptor-dependent long-term potentiation at hippocampal-prefrontal cortex synapses in vivo. *J Neurosci* 106: 1–5, 2000.

195. Gutkin BS, Ermentrout GB. Dynamics of membrane excitability determine interspike interval variability: a link between spike generation mechanisms, and cortical spike train statistics. *Neur Comp* 10: 1047–1065, 1998.

196. Haber SN, Fudge JL, McFarland NR. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J Neurosci* 20: 2369–2382, 2000.

197. Haber SN, Kim KS, Mailly P, Calzavara R. Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. *J Neurosci* 26: 8368–8376, 2006.

198. Hanes DP, Schall JP. Neural control of voluntary movement initiation. *Science* 274: 427–430, 1996.

199. Hanks T, Kiani R, Shadlen MN. A neural mechanism of speed-accuracy tradeoff in macaque area LIP. *eLife* 3: e02260, 2014.

200. Hansel D, Sompolinsky H. Synchronization and computation in a chaotic neural network. *Phys Rev Lett* 68: 718–721, 1992.

201. Hariri AR, Mattay VS, Tessitore A, Kolachana B, Fera F, Goldman D, 2 Egan MF, Weinberger DR. Serotonin transporter genetic variation and the response of the human amygdala. *Science* 297: 400–403, 2002.

202. Harlow HF. The formation of learning sets. *Psychol Rev* 56: 51–65, 1949.

203. Harnett MT, Bernier BE, Ahn KC, Morikawa H. Burst-timing-dependent plasticity of NMDA receptor-mediated transmission in midbrain dopamine neurons. *Neuron* 62: 826–838, 2009.

204. Hart AS, Rutledge RB, Glimcher PW, Phillips PEM. Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *J Neurosci* 34: 698–704, 2014.

205. Hassani OK, Cromwell HC, Schultz W. Influence of expectation of different rewards on behavior-related neuronal activity in the striatum. *J Neurophysiol* 85: 2477–2489, 2001.

206. Hayden BY, Heilbronner SR, Pearson JM, Platt ML. Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *J Neurosci* 31: 4178–4187, 2011.

207. Hayden BY, Heilbronner SR, Platt ML. Ambiguity aversion in rhesus macaques. *Front Neurosci* 4: 166: 1–7, 2010.

208. Hayden BY, Pearson JM, Platt ML. Fictive reward signals in the anterior cingulate cortex. *Science* 324: 948–950, 2009.

209. Hayden BY, Pearson JM, Platt ML. Neuronal basis of sequential foraging decisions in a patchy environment. *Nat Neurosci* 14: 933–939, 2011.

210. Hayden BY, Platt ML. Neurons in anterior cingulate cortex multiplex information about reward and action. *J Neurosci* 30: 3339–3346, 2010.

211. Henry DJ, Greene MA, White FJ. Electrophysiological effects of cocaine in the mesoaccumbens dopamine system: Repeated administration. *J Pharm Exp Ther* 251: 833–839, 1989.

212. Hernández A, Zainos A, Romo R. Temporal evolution of a decision-making process in medial premotor cortex. *Neuron* 33: 959–972, 2002.

213. Hernández-López S, Bargas J, Surmeier DJ, Reyes A, Galarraga E. D1 receptor activation enhances evoked discharge in neostriatal medium spiny neurons by modulating an L-type Ca^{2+} conductance. *J Neurosci* 17: 3334–3342, 1997.

214. Hernández-López S, Tkatch T, Perez-Garci E, Galarraga E, Bargas J, Hamm H, Surmeier DJ. D2 dopamine receptors in striatal medium spiny neurons reduce L-type Ca^{2+} currents and excitability via a novel PLCβ1-IP_3-calcineurin-signaling cascade. *J Neurosci* 20, 8987–8995, 2000.

215. Hikosaka O, Sakamoto M, Usui S. Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J Neurophysiol* 61: 814–832, 1989.

216. Hikosaka K, Watanabe M. Long-range and short-range reward expectancy in the primate orbitofrontal cortex. *Eur J Neurosci* 19: 1046–1054, 2004.

217. Histed MH, Pasupathy A, Miller EK. Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. *Neuron* 63: 244–253, 2009.

218. Ho MY, Mobini S, Chinang TJ, Bradshaw CM, Szabadi E. Theory and method in the quantitative analysis of "impulsive choice" behaviour: implications for psychopharmacology. *Psychopharmacology* 146: 362–372, 1999.

219. Hofstadter DR. *Gödel, Escher, Bach: The Eternal Golden Braid*. New York: Basics Books, 1979, p 711.

220. Holland PC. CS-US interval as a determinant of the form of Pavlovian appetitive conditioned responses. *J Exp Psychol Anim Behav Process* 6: 155–174, 1980.

221. Hollerman JR, Schultz W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat Neurosci* 1: 304–309, 1998.

222. Hollerman JR, Tremblay L, Schultz W. Influence of reward expectation on behavior-related neuronal activity in primate striatum. *J Neurophysiol* 80: 947–963, 1998.

223. Hollis KL, Pharr VL, Dumas MJ, Britton GB, Field J. Classical conditioning provides paternity advantage for territorial male blue gouramis *(Trichogaster trichopterus)*. *J Comp Psychol* 111: 219–225, 1997.

224. Hollon NG, Arnold MM, Gan JO, Walton ME, Phillips PEM. Dopamine-associated cached values are not sufficient as the basis for action selection. *Proc Natl Acad Sci USA* 111: 18357–18362, 2014.

225. Holt CA, Laury SK. Risk aversion and incentive effects. *Am Econ Rev* 92: 1644–1655, 2002.

226. Hong S, Jhou TC, Smith M, Saleem KS, Hikosaka O. Negative reward signals from the lateral habenula to dopamine neurons are mediated by rostromedial tegmental nucleus in primates. *J Neurosci* 31: 11457–11471, 2011.

227. Hong S, Hikosaka O. The globus pallidus sends reward-related signals to the lateral habenula. *Neuron* 60, 720–729, 2008.

228. Hong S, Hikosaka O. Pedunculopontine tegmental nucleus neurons provide reward, sensorimotor, and alerting signals to midbrain dopamine neurons. *Neuroscience* 282: 139–155, 2014.

229. Horvitz JC, Stewart Jacobs BL. Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. *Brain Res* 759: 251–258, 1997.

230. Horwitz GD, Batista AP, Newsome WT. Representation of an abstract perceptual decision in macaque superior colliculus. *J Neurophysiol* 91: 2281–2296, 2004.

231. Hosokawa T, Kato K, Inoue M, Mikami A. Neurons in the orbitofrontal cortex code both visual shapes and reward types. *NeuroReport* 15: 1493–1496, 2004.

232. Hosokawa T, Kato K, Inoue M, Mikami A. Neurons in the macaque orbitofrontal cortex code relative preference of both rewarding and aversive outcomes. *Neurosci Res* 57: 434–445, 2007.

233. Hosokawa T, Kennerley SW, Sloan J, Wallis JD. Single-neuron mechanisms underlying cost-benefit analysis in frontal cortex. *J Neurosci* 33: 17385–17397, 2013.

234. Hosokawa T, Watanabe M. Prefrontal neurons represent winning and losing during competitive video shooting games between monkeys. *J Neurosci* 32: 7662–7671, 2012.

235. Hosoya T, Baccus SA, Meister M. Dynamic predictive coding by the retina. *Nature* 436: 71–77, 2005.

236. Houk JC, Adams JL, Barto AG. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, Beiser DG. Cambridge, MA: MIT Press, 1995, p. 249–270.

237. Howe MW, Tierney PL, Sandberg SG, Phillips PEM, Graybiel AM. Prolonged dopamine signaling in striatum signals proximity and value of distant rewards. *Nature* 500: 575–579, 2013.

238. Huang CF, Litzenberger RH. *Foundations for Financial Economics*. Upper Saddle River, NJ: Prentice-Hall, 1988.

239. Hrupka BJ, Lin YM, Gietzen DW, Rogers QR. Small changes in essential amino acid concentrations alter diet selection in amino acid-deficient rats. *J Nutr* 127: 777–784, 1997.

240. Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF. Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310: 1680–1683, 2005.

241. Hsu M, Krajbich I, Zhao C, Camerer CF. Neural response to reward anticipation under risk is nonlinear in probabilities. *J Neurosci* 29: 2237–2231, 2009.

242. Hull CL. *Principles of Behavior*. New York: Appleton-Century-Crofts, 1943.

243. Ikeda T, Hikosaka O. Reward-dependent gain and bias of visual responses in primate superior colliculus. *Neuron* 39: 693–700, 2003.

244. Ilango A, Kesner AJ, Keller KL, Stuber GD, Bonci A, Ikemoto S. Similar roles of substantia nigra and ventral tegmental dopamine neurons in reward and aversion. *J Neurosci* 34: 817–822, 2014.

245. Ito M, Doya K. Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *J Neurosci* 29: 9861–9874, 2009.

246. Ito S, Stuphorn V, Brown JW, Schall JD. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* 302: 120–122, 2003.

247. Izhikevich EM. Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cereb Cortex* 17: 2443–2452, 2007.

248. Janssen P, Shadlen MN. A representation of the hazard rate of elapsed time in macaque area LIP. *Nat Neurosci* 8: 234–241, 2005.

249. Jhou TC, Fields HL, Baxter MB, Saper CB, Holland PC. The rostromedial tegmental nucleus (RMTg), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. *Neuron* 61: 786–800, 2009.

250. Ji H, Shepard PD. Lateral habenula stimulation inhibits rat midbrain dopamine neurons through a GABAA receptor-mediated mechanism. *J Neurosci* 27: 6923–6930, 2007.

251. Jimenez-Castellanos J, Graybiel AM. Subdivisions of the dopamine-containing A8-A9-A10 complex identified by their differential mesostriatal innervation of striosomes and extrastriosomal matrix. *Neuroscience* 23: 223–242, 1987.

252. Jimenez-Castellanos J, Graybiel AM. Evidence that histochemically distinct zones of the primate substantia nigra pars compacta are related to patterned distributions of nigrostriatal projection neurons and striatonigral fibers. *Exp Brain Res* 74: 227–238, 1989.

253. Jog MS, Kubota Y, Connolly CI, Hillegaart V, Graybiel AM. Building neural representations of habits. *Science* 286: 1745–1749, 1999.

254. Johnson JG, Busemeyer JR. A dynamic, stochastic, computational model of preference reversal phenomena. *Psychol Rev* 112: 841–861, 2005.

255. Joshua M, Adler A, Mitelman R, Vaadia E, Bergman H. Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. *J Neurosci* 28: 11673–11684, 2008.

256. Joshua M, Adler A, Prut Y, Vaadia E, Wickens JR, Hagai Bergman H. Synchronization of midbrain dopaminergic neurons is enhanced by rewarding events. *Neuron* 62: 695–704, 2009.

257. Kagel JH, Battalio RC, Green L. *Economic Choice Theory: An Experimental Analysis of Animal Behavior*. Cambridge, UK: Cambridge Univ. Press, 1995.

258. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica* 47: 263–291, 1979.

259. Kahneman D, Wakker PP, Sarin R. Back to Bentham? Explorations of experienced utility. *Q J Econ* 112: 375–405, 1997.

260. Kawagoe R, Takikawa Y, Hikosaka O. Expectation of reward modulates cognitive signals in the basal ganglia. *Nat Neurosci* 1: 411–416, 1998.

261. Kennerley SW, Behrens TEJ, Wallis JD. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat Neurosci* 14: 1581–1589, 2011.

262. Kennerley SW, Wallis JD. Reward-dependent modulation of working memory in lateral prefrontal cortex. *J Neurosci* 29: 3259–3270, 2009.

263. Kennerley SW, Wallis JD. Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. *Eur J Neurosci* 29: 2061–2073, 2009.

264. Kepecs A, Uchida N, Zariwala H, Mainen ZF. Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455: 227–231, 2008.

265. Kerr JN, Wickens JR. Dopamine D-1/D-5 receptor activation is required for long-term potentiation in the rat neostriatum in vitro. *J Neurophysiol* 85: 117–124, 2001.

266. Kettner RE, Mahamud S, Leung HC, Sitkoff N, Houk JC, Peterson BW, Barto AG. Prediction of complex two-dimensional trajectories by a cerebellar model of smooth pursuit eye movements. *J Neurophysiol* 77: 2115–2130, 1997.

267. Khamassi M, Quilodran R, Pierre Enel P, Dominey PF, Procyk E. Behavioral regulation and the modulation of information coding in the lateral prefrontal and cingulate cortex. *Cereb Cortex*. In press.

268. Kheramin S, Body S, Ho MY, Velazquez-Martinez DN, Bradshaw CM, Szabadi E, Deakin JFW, Anderson IM. Effects of orbital prefrontal cortex dopamine depletion on inter-temporal choice: a quantitative analysis. *Psychopharmacology* 175: 206–214, 2004.

269. Kiani R, Cueva CJ, Reppas JB, Newsome WT. Dynamics of neural population responses in prefrontal cortex indicate changes of mind on single trials. *Curr Biol* 24: 1542–1547, 2014.

270. Kiani R, Hanks TD, Shadlen MN. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J Neurosci* 28: 3017–3029, 2008.

271. Kim KM, Baratta MV, Yang A, Lee D, Boyden ES, Fiorillo CD. Optogenetic mimicry of the transient activation of dopamine neurons by natural reward is sufficient for operant reinforcement. *PLoS One* 7: e33612, 2012.

272. Kim S, Hwang J, Lee D. Prefrontal coding of temporally discounted values during intertemporal choice. *Neuron* 59: 161–172, 2008.

273. Kim JJ, Krupa DJ, Thompson RF. Inhibitory cerebello-olivary projections and blocking effect in classical conditioning. *Science* 279: 570–573, 1998.

274. Kim JN, Shadlen MN. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nat Neurosci* 2: 176–185, 1999.

275. Kim H, Sul JH, Huh N, Lee D, Jung MW. Role of striatum in updating values of chosen actions. *J Neurosci* 29: 14701–14712, 2009.

276. Kimura M, Rajkowski J, Evarts E. Tonically discharging putamen neurons exhibit set-dependent responses. *Proc Natl Acad Sci USA* 81: 4998–5001, 1984.

277. Kirby KN, Marakovic NN. Delay-discounting probabilistic rewards: rates decrease as amounts increase. *Psychonom Bull Rev* 3: 100–104, 1996.

278. Kitazawa S, Kimura T, Yin PB. Cerebellar complex spikes encode both destinations and errors in arm movement. *Nature* 392: 494–497, 1998.

279. Klein JT, Deaner RO, Platt ML. Neural correlates of social target value in macaque parietal cortex. *Curr Biol* 18: 419–424, 2008.

280. Knutson B, Adams CM, Fong GW, Hommer D. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci* 21: 1–5, 2001.

281. Kobayashi Y, Inoue Y, Yamamoto M, Isa T, Aizawa H. Contribution of pedunculopontine tegmental nucleus neurons to performance of visually guided saccade tasks in monkeys. *J Neurophysiol* 88: 715–731, 2002.

282. Kobayashi S, Lauwereyns J, Koizumi M, Sakagami M, Hikosaka O. Influence of reward expectation on visuospatial processing in macaque lateral prefrontal cortex. *J Neurophysiol* 87: 1488–1498, 2002.

283. Kobayashi S, Nomoto K, Watanabe M, Hikosaka O, Schultz W, Sakagami M. Influences of rewarding and aversive outcomes on activity in macaque lateral prefrontal cortex. *Neuron* 51: 861–870, 2006.

284. Kobayashi S, Pinto de Carvalho O, Schultz W. Adaptation of reward sensitivity in orbitofrontal neurons. *J Neurosci* 30: 534–544, 2010.

285. Kobayashi S, Schultz W. Influence of reward delays on responses of dopamine neurons. *J Neurosci* 28: 7837–7846, 2008.

286. Kobayashi S, Schultz W. Reward contexts extend dopamine signals to unrewarded stimuli. *Curr Biol* 24: 56–62, 2014.

287. Koch C. *Biophysics of Computation*. New York: Oxford Univ. Press, 1999.

288. Komura Y, Nikkuni A, Hirashima N, Uetake T, Miyamoto A. Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nat Neurosci* 16: 749–755, 2013.

289. Konorski J. *Integrative Activity of the bBrain*. Chicago: Univ. of Chicago Press, 1967.

290. Köszegi B, Rabin M. A model of reference-dependent preferences. *Q J Econ* 121: 1133–1165, 2006.

291. Krajbich I, Armel C, Rangel R. Visual fixations and the computation and comparison of value in simple choice. *Nat Neurosci* 13: 1292–1298, 2010.

292. Krauzlis RJ, Basso MA, Wurtz RH. Shared motor error for multiple eye movements. *Science* 276: 1693–1695, 1997.

293. Kravitz AV, Tye LD, Kreitzer AC. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nat Neurosci* 15: 816–818, 2012.

294. Kreitzer AC, Malenka RC. Endocannabinoid-mediated rescue of striatal LTD and motor deficits in Parkinson's disease models. *Nature* 445: 643–647, 2007.

295. Kreps DM. *A Course in Microeconomic Theory*. Harlow: Pearson Education, 1990.

296. Kreps DM, Porteus E. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* 46: 185–200, 1978.

297. Kuhn A, Aertsen A, Rotter S. Neuronal integration of synaptic input in the fluctuation-driven regime. *J Neurosci* 24, 2345–2356, 2004.

298. Kurata K, Wise SP. Premotor cortex of rhesus monkeys: set-related activity during two conditional motor tasks. *Exp Brain Res* 69: 327–343, 1988.

299. Kurata K, Wise SP. Premotor and supplementary motor cortex in rhesus monkeys: neuronal activity during externally- and internally-instructed motor tasks. *Exp Brain Res* 72: 237–248, 1988.

300. Lak A, Arabzadeh E, Harris JA, Diamond ME. Correlated physiological and perceptual effects of noise in a tactile stimulus. *Proc Natl Acad Sci USA* 107: 7981–7986, 2010.

301. Lak A, Stauffer WR, Schultz W. Dopamine prediction error responses integrate subjective value from different reward dimensions. *Proc Natl Acad Sci USA* 111: 2343–2348, 2014.

302. Laming DRJ. *Information Theory of Choice Reaction Time*. New York: Wiley, 1968.

303. Lammel S, Lim BK, Ran C, Huang KW, Betley MJ, Tye KM, Deisseroth K, Malenka RC. Input-specific control of reward and aversion in the ventral tegmental area. *Nature* 491: 212–217, 2012.

304. Lammel S, Steinberg E, Földy C, Wall NR, Beier K, Luo L, Malenka RC. Diversity of transgenic mouse models for selective Ttrgeting of midbrain dopamine neurons. *Neuron* 85: 429–438, 2015.

305. Lattimore PK, Baker JR, Witte AD. The influence of probability on risky choice: a parametric examination. *J Econ Behav Organ* 17: 377–400, 1992.

306. Lau B, Glimcher PW. Value representations in the primate striatum during matching behavior. *Neuron* 58: 451–463, 2008.

307. Laughlin S. A simple coding procedure enhances a neuron's information capacity. *Z Naturforsch* 36: 910–912, 1981.

308. Lauwereyns J, Takikawa Y, Kawagoe R, Kobayashi S, Koizumi M, Coe B, Sakagami M, Hikosaka O. Feature-based anticipation of cues that predict reward in monkey caudate nucleus. *Neuron* 33: 463–473, 2002.

309. Leathers ML, Olson CR. In monkeys making value-based decisions, LIP neurons encode cue salience and not action value. *Science* 338: 132–135, 2012.

310. Lecourtier L, DeFrancesco A, Moghaddam B. Differential tonic influence of lateral habenula on prefrontal cortex and nucleus accumbens dopamine release. *Eur J Neurosci* 27: 1755–1762, 2008.

311. Lee IH, Assad JA. Putaminal activity for simple reactions or self-timed movements. *J Neurophysiol* 89: 2528–2537, 2003.

312. Lemus L, Hernández A, Luna R, Zainos A, Nácher V, Romo R. Neural correlates of a postponed decision report. *Proc Natl Acad Sci USA* 104: 17174–17179, 2007.

313. Leon MI, Shadlen MN. Effect of expected reward magnitude on the responses of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* 24: 415–425, 1999.

314. Leung PMB, Rogers QR. Importance of prepyriform cortex in food-intake response of rats to amino acids. *Am J Physiol* 221: 929–935, 1971.

315. Levy H, Markowitz HM. Approximating expected utility by a function of mean and variance. *Am Econ Rev* 69: 308–317, 1979.

316. Libet B, Gleason CA, Wright EW, Pearl DK. Time of conscious intention to act in relation to onset of cerebral activities (readiness-potential): the unconscious initiation of a freely voluntary act. *Brain* 106: 623–642, 1983.

317. Lishman WA. *Organic Psychiatry*. Oxford: Blackwell, 1998.

318. Liu QS, Pu L, Poo MM. Repeated cocaine exposure in vivo facilitates LTP induction in midbrain dopamine neurons. *Nature* 437: 1027–1031, 2005.

319. Liu Z, Richmond BJ. Response differences in monkey TE and perirhinal cortex: stimulus association related to reward schedules. *J Neurophysiol* 83: 1677–1692, 2000.

320. Livio M. *The Golden Ratio: The Story of PHI, the World's Most Astonishing Number*. New York: Random House, 2002.

321. Ljungberg T, Apicella P, Schultz W. Responses of monkey midbrain dopamine neurons during delayed alternation performance. *Brain Res* 586: 337–341, 1991.

322. Ljungberg T, Apicella P, Schultz W. Responses of monkey dopamine neurons during learning of behavioral reactions. *J Neurophysiol* 67: 145–163, 1992.

323. Lo CC, Wang XJ. Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat Neurosci* 9: 956–963, 2006.

324. Loewenstein G, Thompson L, Bazerman MH. Social utility and decision making in interpersonal contexts. *J Personality Soc Psych* 57: 426–441, 1989.

325. Loewenstein G, Prelec D. Anomalies in intertemporal choice: evidence and an interpretation. *Q J Econ* 107: 573–597, 1992.

326. Leszczuk MH, Flaherty CF. Lesions of nucleus accumbens reduce instrumental but not consummatory negative contrast in rats. *Behav Brain Res* 116: 61–79, 2000.

327. Logan GD, Cowan WB. On the ability to inhibit thought and action: a theory of an act of control. *Psychol Rev* 91: 295–327, 1984.

328. Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412: 150–157, 2001.

329. Lohrenz T, McCabe K, Camerer CF, Montague PR. Neural signature of fictive learning signals in a sequential investment task. *Proc Natl Acad Sci USA* 104: 9493–9498, 2007.

330. Longtin A. Stochastic resonance in neuron models. *J Stat Phys* 70: 309, 1993.

331. Louie K, Glimcher PW. Separating value from choice: delay discounting activity in the lateral intraparietal area. *J Neurosci* 30: 5498–5507, 2010.

332. Louie K, Grattan LE, Glimcher PW. Reward value-based gain control: divisive normalization in parietal cortex. *J Neurosci* 31: 10627–10639, 2011.

333. Luce RD. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley, 1959.

334. Machens CK, Romo R, Brody CD. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* 307: 1121–1124, 2005.

335. Mackintosh NJ. *The Psychology of Animal Learning*. London: Academic Press, 1974.

336. Mackintosh NJ. A theory of attention: variations in the associability of stimulus with reinforcement. *Psychol Rev* 82: 276–298, 1975.

337. Maimon G, Assad JA. Parietal area 5 and the initiation of self-timed movements versus simple reactions. *J Neurosci* 26: 2487–2498, 2006.

338. Markowitsch HJ, Pritzel M. Reward-related neurons in cat association cortex. *Brain Res* 111: 185–188, 1976.

339. Markowitz H. The utility of wealth. *J Polit Econ* 6: 151–158, 1952.

340. Martínez-García M, Rolls ET, Deco G, Romo R. Neural and computational mechanisms of postponed decisions. *Proc Natl Acad Sci* 108: 11626–11631, 2011.

341. Mas-Colell A, Whinston M, Green J. *Microeconomic Theory*. New York: Oxford Univ. Press, 1995.

342. Matsuda W, Furuta T, Nakamura KC, Hioki H, Fujiyama F, Arai R, Kaneko T. Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. *J Neurosci* 29: 444–453, 2009.

343. Matsuda Y, Marzo A, Otani S. The presence of background dopamine signal converts long-term synaptic depression to potentiation in rat prefrontal cortex. *J Neurosci* 26: 4803–4810, 2006.

344. Matsumoto M, Hikosaka O. Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* 447: 1111–1115, 2007.

345. Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctively convey positive and negative motivational signals. *Nature* 459: 837–841, 2009.

346. Matsumoto M, Hikosaka O. Representation of negative motivational value in the primate lateral habenula. *Nat Neurosci* 12: 77–84, 2009.

347. Matsumoto M, Matsumoto K, Abe H, Tanaka K. Medial prefrontal cell activity signaling prediction errors of action values. *Nat Neurosci* 10: 647–656, 2007.

348. Matsumoto K, Suzuki W, Tanaka K. Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* 301: 229–232, 2003.

349. Matsumoto M, Takada M. Distinct representations of cognitive and motivational signals in midbrain dopamine neurons. *Neuron* 79: 1011–1024, 2013.

350. Mauritz KH, Wise SP. Premotor cortex of the rhesus monkey: neuronal activity in anticipation of predictable environmental events. *Exp Brain Res* 61: 229–244, 1986.

351. McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38: 339–346, 2003.

352. McClure SM, Laibson DI, Loewenstein G, Cohen JD. Separate neural systems value immediate and delayed monetary rewards. *Science* 306: 503–507, 2004.

353. McClure SM, Li J, Tomlin D, Cypert KS, Montague LM, Montague PR. Neural correlates of behavioral preference for culturally familiar drinks. *Neuron* 44: 379–387, 2004.

354. McCoy AN, Crowley JC, Haghighian G, Dean HL, Platt ML. Saccade reward signals in posterior cingulate cortex. *Neuron* 40: 1031–1040, 2003.

355. McCoy AN, Platt ML. Risk-sensitive neurons in macaque posterior cingulate cortex. *Nat Neurosci* 8: 1220–1227, 2005.

356. McEchron MD, Bouwmeester H, Tseng W, Weiss C, Disterhoft JF. Hippocampectomy disrupts auditory trace fear conditioning and contextual fear conditioning in the rat. *Hippocampus* 8: 638–646, 1998.

357. Medina JF, Nores WL, Mauk MD. Inhibition of climbing fibers is a signal for the extinction of conditioned eyelid responses. *Nature* 416: 330–333, 2002.

358. Meijer JH, Robbers Y. Wheel running in the wild. *Proc R Soc B* 281: 2014.330–0210, 2014.

359. Melis AP, Hare B, Tomasello M. Engineering cooperation in chimpanzees: tolerance constraints on cooperation. *Anim Behav* 72: 275–286, 2006.

360. Merten K, Nieder A. Active encoding of decisions about stimulus absence in primate prefrontal cortex neurons. *Proc Natl Acad Sci USA* 109: 6289–6294, 2012.

361. Merten K, Nieder A. Comparison of abstract decision encoding in the monkey prefrontal cortex, the presupplementary, and cingulate motor areas. *J Neurophysiol* 110: 19–32, 2013.

362. Mileykovskiy B, Morales M. Duration of inhibition of ventral tegmental area dopamine neurons encodes a level of conditioned fear. *J Neurosci* 31: 7471–7476, 2011.

363. Mill JS. *Utilitarianism.* London: Parker, Son and Bourn, 1863.

364. Miller LA. Cognitive risk-taking after frontal or temporal lobectomy. I: Synthesis of fragmented visual information. *Neuropsychologia* 23: 359–369, 1985.

365. Mirenowicz J, Schultz W. Importance of unpredictability for reward responses in primate dopamine neurons. *J Neurophysiol* 72: 1024–1027, 1994.

366. Mirenowicz J, Schultz W. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* 379: 449–451, 1996.

367. Mitchell DS, Gormezano I. Effects of water deprivation on classical appetitive conditioning of the rabbit's jaw movement response. *Learn Motivat* 1: 199–206, 1970.

368. Mink JW. The basal ganglia: focused selection and inhibition of competing motor programs. *Prog Neurobiol* 50: 381–425, 1996.

369. Mizuhiki T, Richmond BJ, Shidara M. Encoding of reward expectation by monkey anterior insular neurons. *J Neurophysiol* 107: 2996–3007, 2012.

370. Mobini S, Body S, Ho MY, Bradshaw CM, Szabadi E, Deakin JFW, Anderson IM. Effects of lesions of the orbitofrontal cortex on sensitivity to delayed and probabilistic reinforcement. *Psychopharmacology* 160: 290–298, 2002.

371. Mogami T, Tanaka K. Reward association affects neuronal responses to visual stimuli in macaque TE and perirhinal cortices. *J Neurosci* 26: 6761–6770, 2006.

372. Monosov IE, Hikosaka O. Regionally distinct processing of rewards and punishments by the primate ventromedial prefrontal cortex. *J Neurosci* 32: 10318–10330, 2012.

373. Monosov IE, Hikosaka O. Selective and graded coding of reward uncertainty by neurons in the primate anterodorsal septal region. *Nat Neurosci* 16: 756–762, 2013.

374. Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16: 1936–1947, 1996.

375. Montague PR, Sejnowski TJ. The predictive brain: temporal coincidence and temporal order in synaptic learning mechanisms. *Learn Mem* 1: 1–33, 1994.

376. Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H. Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron* 43: 133–143, 2004.

377. Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H. Midbrain dopamine neurons encode decisions for future action. *Nat Neurosci* 9: 1057–1063, 2006.

378. Morrison SE, Saez A, Lau B, Salzman CD. Different time courses for learning-related changes in amygdala and orbitofrontal cortex. *Neuron* 71, 1127–1140, 2011.

379. Munoz DP, Wurtz RH. Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. *J Neurophysiol* 73: 2313–2333, 1995.

380. Murakami M, Vicente MI, Costa GM, Mainen ZF. Neuronal antecedents of self-initiated actions in secondary motor cortex. *Nat Neurosci* 17: 1574–1582, 2014.

381. Musallam S, Corneil BD, Greger B, Scherberger H, Andersen RA. Cognitive control signals for neural prosthetics. *Science* 305: 258–262, 2004.

382. Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O. Dopamine neurons can represent context-dependent prediction error. *Neuron* 41: 269–280, 2004.

383. Nakamura K, Mikami A, Kubota K. Activity of single neurons in the monkey amygdala during performance of a visual discrimination task. *J Neurophysiol* 67: 1447–1463, 1992.

384. Nassar MR, Wilson RC, Heasly B, Gold JI. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J Neurosci* 30: 12366–12378, 2010.

385. Newsome WT, Britten KH, Movshon JA. Neuronal correlates of a perceptual decision. *Nature* 341: 52–54, 1989.

386. Niki H, Watanabe M. Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Res* 171: 213–224, 1979.

387. Nishijo H, Ono T, Nishino H. Single neuron responses in amygdala of alert monkey during complex sensory stimulation with affective significance. *J Neurosci* 8: 3570–3583, 1988.

388. Niv Y, Duff MO, Dayan P. Dopamine, uncertainty and TD learning. *Behav Brain Func* 1:6, 2005.

389. Nomoto K, Schultz W, Watanabe T, Sakagami M. Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. *J Neurosci* 30: 10692–10702, 2010.

390. O'Doherty J, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related lNiv earning in the human brain. *Neuron* 28: 329–337, 2003.

391. O'Neill M, Schultz W. Coding of reward risk distinct from reward value by orbitofrontal neurons. *Neuron* 68: 789–800, 2010.

392. O'Neill M, Schultz W. Risk prediction error coding in orbitofrontal neurons. *J Neurosci* 33: 15810–15814, 2013.

393. Ogawa M, van der Meer MAA, Esber GR, Cerri DH, Stalnaker TA, Schoenbaum G. Risk-responsive orbitofrontal neurons track acquired salience. *Neuron* 77: 251–258, 2013.

394. Ohyama K, Sugase-Miyamoto Y, Matsumoto N, Shidara M, Sato C. Stimulus-related activity during conditional associations in monkey perirhinal cortex neurons depends on upcoming reward outcome. *J Neurosci* 32: 17407–17419, 2012.

395. Ojakangas CL, Ebner TJ. Purkinje cell complex and simple spike changes during a voluntary arm movement learning task in the monkey. *J Neurophysiol* 68: 2222–2236, 1992.

396. Okada KI, Toyama K, Inoue Y, Isa T, Kobayashi Y. Different pedunculopontine tegmental neurons signal predicted and actual task rewards. *J Neurosci* 29: 4858–4870, 2009.

397. Okano K, Tanji J. Neuronal activities in the primate motor fields of the agranular frontal cortex preceding visually triggered and self-paced movement. *Exp Brain Res* 66: 155–166, 1987.

398. Olds J, Milner P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J Comp Physiol Psychol* 47: 419–427, 1954.

399. Ostlund SB, Balleine BW. Differential involvement of the basolateral amygdala and mediodorsal thalamus in instrumental action selection. *J Neurosci* 28: 4398–4405, 2008.

400. Otani S, Blond O, Desce JM, Crepel F. Dopamine facilitates long-term depression of glutamatergic transmission in rat prefrontal cortex. *Neuroscience* 85: 669–676, 1998.

401. Otmakhova NA, Lisman JE. D1/D5 dopamine receptor activation increases the magnitude of early long-term potentiation at CA1 hippocampal synapses. *J Neurosci* 16: 7478–7486, 1996.

402. Oyama K, Hernádi I, Iijima T, Tsutsui KI. Reward prediction error coding in dorsal striatal neurons. *J Neurosci* 30: 11447–11457, 2010.

403. Padoa-Schioppa C. Range-adapting representation of economic value in the orbitofrontal cortex. *J Neurosci* 29: 14004–14014, 2009.

404. Padoa-Schioppa C. Neuronal origins of choice variability in economic decisions. *Neuron* 80: 1322–1336, 2013.

405. Padoa-Schioppa C, Assad JA. Neurons in the orbitofrontal cortex encode economic value. *Nature* 441: 223–226, 2006.

406. Padoa-Schioppa C, Assad JA. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat Neurosci* 11: 95–102, 2008.

407. Paladini CA, Celada P, Tepper JM. Striatal, pallidal, and pars reticulata evoked inhibition of nigrostriatal dopaminergic neurons is mediated by GABAa receptors in vivo. *Neuroscience* 89: 799–812, 1998.

408. Pan WX, Brown J, Dudman JT. Neural signals of extinction in the inhibitory microcircuit of the ventral midbrain. *Nat Neurosci* 16: 71–78, 2013.

409. Pan X, Fan H, Sawa K, Tsuda I, Tsukada M, Sakagami M. Reward inference by primate prefrontal and striatal neurons. *J Neurosci* 34: 1380–1396, 2014.

410. Pan WX, Hyland BI. Pedunculopontine tegmental nucleus controls conditioned responses of midbrain dopamine neurons in behaving rats. *J Neurosci* 25: 4725–4732, 2005.

411. Pan X, Sawa K, Tsuda I, Tsukada M, Sakagami M. Reward prediction based on stimulus categorization in primate lateral prefrontal cortex. *Nat Neurosci* 11: 703–712, 2008.

412. Pan WX, Schmidt R, Wickens JR, Hyland BI. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J Neurosci* 25: 6235–6242, 2005.

413. Parker JG, Beutler LR, Palmiter RD. The contribution of NMDA receptor signaling in the corticobasal ganglia reward network to appetitive Pavlovian learning. *J Neurosci* 31: 11362–11369, 2011.

414. Parker JG, Wanat MJ, Soden ME, Ahmad K, Zweifel LS, Bamford NS, Palmiter RD. Attenuating $GABA_A$ receptor signaling in dopamine neurons selectively enhances reward learning and alters risk preference in mice. *J Neurosci* 31: 17103–17112, 2011.

415. Parker JG, Zweifel LS, Clark JJ, Evans SB, Phillips PEM, Palmiter RD. Absence of NMDA receptors in dopamine neurons attenuates dopamine release but not conditioned approach during Pavlovian conditioning. *Proc Nat Acad Sci USA* 107: 13491–13496, 2010.

416. Parthasarathy HB, Schall JD, Graybiel AM. Distributed but convergent ordering of corticostriatal projections: analysis of the frontal eye field and the supplementary eye field in the macaque monkey. *J Neurosci* 12: 4468–4488, 1992.

417. Pascal B. *Pensées 1658–1662*, translated by Ariew R. Indianapolis: Hackett, 2004.

418. Pasquereau B, Nadjar A, Arkadir D, Bezard E, Goillandeau M, Bioulac B, Gross CE, Boraud T. Shaping of motor responses by incentive values through the basal ganglia. *J Neurosci* 27: 1176–1183, 2007.

419. Pasquereau B, Turner RS. Limited encoding of effort by dopamine neurons in a cost-benefit trade-off task. *J Neurosci* 33: 8288–8300, 2013.

420. Pastor-Bernier A, Cisek P. Neural correlates of biased competition in premotor cortex. *J Neurosci* 31: 7083–7088, 2011.

421. Pasupathy A, Miller EK. Different time courses of learning-related activity in the prefrontal cortex and striaum. *Nature* 433: 873–876, 2005.

422. Paton JJ, Belova MA, Morrison SE, Salzman CD. The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* 439: 865–870, 2006.

423. Pavlov PI. *Conditioned Reflexes*. London: Oxford Univ. Press, 1927.

424. Pawlak V, Kerr JND. Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. *J Neurosci* 28: 2435–2446, 2008.

425. Pearce JM, Hall G. A model for Pavlovian conditioning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev* 87: 532–552, 1980.

426. Pearson JM, PLatt ML. Dopamine: burning the candle at both ends. *Neuron* 79: 831–833, 2013.

427. Peck CJ, Jangraw DC, Suzuki M, Efem R, Gottlieb J. Reward modulates attention independently of action value in posterior parietal cortex. *J Neurosci* 29: 11182–11191, 2009.

428. Peck CJ, Lau B, Salzman CD. The primate amygdala combines information about space and value. *Nat Neurosci* 16: 340–348, 2013.

429. Penrose R. *The Emperor's New Mind*. Oxford, UK: Oxford Univ. Press, 1989.

430. Percheron G, Yelnik J, Francois C. A Golgi analysis of the primate globus pallidus. III. Spatial organization of the striopallidal complex. *J Comp Neurol* 227: 214–227, 1984.

431. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442: 1042–1045, 2006.

432. Plassmann H, O'Doherty J, Shiv B, Rangel A. Marketing actions can modulate neural representations of experienced pleasantness. *Proc Natl Acad Sci USA* 105: 1050–1054, 2009.

433. Platt ML, Glimcher PW. Neural correlates of decision variables in parietal cortex. *Nature* 400: 233–238, 1999.

434. Pooresmaeili A, Poort J, Roelfsema PR. Simultaneous selection by object-based attention in visual and frontal cortex. *Proc Natl Acad Sci USA* 111: 6467–6472, 2014.

435. Prelec D. The probability weighting function. *Econometrica* 66: 497–527, 1998.

436. Prelec D, Loewenstein G. Decision making over time and under uncertainty: a common approach. *Management Sci* 37: 770–786, 1991.

437. Preuschoff K, Bossaerts P, Quartz SR. Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* 51: 381–390, 2006.

438. Preuschoff Bossaerts P. Adding prediction risk to the theory of reward learning. *Ann NY Acad Sci* 1104: 135–146, 2007.

439. Prévost C, Pessiglione M, Météreau E, Cléry-Melin ML, Dreher JC. Separate valuation subsystems for delay and effort decision costs. *J Neurosci* 30: 14080–14090, 2010.

440. Puig MV, Miller EK. The role of prefrontal dopamine D1 receptors in the neural mechanisms of associative learning. *Neuron* 74: 874–886, 2012.

441. Purcell BA, Heitz RP, Cohen JY, Schall JD, Logan GD, Palmeri TJ. Neurally constrained modeling of perceptual decision making. *Psychol Rev* 117: 1113–1143, 2010.

442. Rabin M. Incorporating fairness into game theory and economics. *Am Econ Rev* 83: 1281–1302, 1993.

443. Raby CR, Alexis DM, Dickinson A, Clayton NS. Planning for the future by western scrub-jays. *Nature* 445: 919–921, 2007.

444. Raghuraman AP, Padoa-Schioppa C. Integration of multiple determinants in the neuronal computation of economic values. *J Neurosci* 34: 11583–11603, 2014.

445. Rahman S, Sahakian BJ, Hodges JR, Rogers RD, Robbins TW. Specific cognitive deficits in mild frontal variant frontotemporal dementia. *Brain* 122: 1469–1493, 1999.

446. Rao RPN, Sejnowski TJ. Self-organizing neural systems based on predictive learning. *Phil Trans R Soc A* 361: 1149–1175, 2003.

447. Ratcliff R. A theory of memory retrieval. *Psychol Rev* 83: 59–108, 1978.

448. Ratcliff R, Cherian A, Segraves M. A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *J Neurophysiol* 90: 1392–1407, 2003.

449. Ratcliff R, Frank MJ. Reinforcement-based decision making in corticostriatal circuits: mutual constraints by neurocomputational and diffusion models. *Neural Comput* 24: 1186–1229, 2012.

450. Ratcliff R, Hasegawa YT, Hasegawa RP, Smith PL, Segraves MA. Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *J Neurophysiol* 97: 1756–1774, 2007.

451. Ravel S, Legallet E, Apicella P. Tonically active neurons in the monkey striatum do not preferentially respond to appetitive stimuli. *Exp Brain Res* 128: 531–534, 1999.

452. Ravel S, Legallet E, Apicella P. Responses of tonically active neurons in the monkey striatum discriminate between motivationally opposing stimuli. *J Neurosci* 23: 8489–8497, 2003.

453. Ravel S, Richmond BJ. Dopamine neuronal responses in monkeys performing visually cued reward schedules. *Eur J Neurosci* 24: 277–290, 2006.

454. Reddi BA, Asrress KN, Carpenter RH. Accuracy, information, and response time in a saccadic decision task. *J Neurophysiol* 90: 3538–3546, 2003.

455. Redgrave P, Prescott TJ, Gurney K. Is the short-latency dopamine response too short to signal reward? *Trends Neurosci* 22: 146–151, 1999.

456. Redgrave P, Gurney K. The short-latency dopamine signal: a role in discovering novel actions? *Nat Rev Neurosci* 7: 967–975, 2006.

457. Redish AD. Addiction as a computational process gone awry. *Science* 306: 1944–1947, 2004.

458. Reed P, Mitchell C, Nokes T. Intrinsic reinforcing properties of putatively neutral stimuli in an instrumental two-lever discrimination task. *Anim Learn Behav* 24: 38–45, 1996.

459. Rescorla RA. Pavlovian conditioning and its proper control procedures. *Psychol Rev* 74: 71–80, 1967.

460. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: *Classical Conditioning II: Current Research and Theory*, edited by Black AH, Prokasy WF. New York: Appleton Century Crofts, 1972, p. 64–99.

461. Reynolds JNJ, Hyland BI, Wickens JR. A cellular mechanism of reward-related learning. *Nature* 413: 67–70, 2001.

462. Richards JB, Mitchell SH, de Wit H, Seiden LS. Determination of discount functions in rats with an adjusting-amount procedure. *J Exp Anal Behav* 67: 353–366, 1997.

463. Richards CD, Shiroyama T, Kitai ST. Electrophysiological and immunocytochemical characterization of GABA and dopamine neurons in the substantia nigra of the rat. *Neuroscience* 80: 545–557, 1997.

464. Richardson KA, Aston-Jones G. Lateral hypothalamic orexin/hypocretin neurons that project to ventral tegmental area are differentially activated with morphine preference. *J Neurosci* 32: 3809–3817, 2012.

465. Richardson MJE, Gerstner W. Statistics of subthreshold neuronal voltage fluctuations due to conductance based synaptic shot noise. *Chaos* 16: 026106, 2006.

466. Richardson RT, DeLong MR. Context-dependent responses of primate nucleus basalis neurons in a go/no-go task. *J Neurosci* 10: 2528–2540, 1990.

467. Riehle A, Requin J. Monkey primary motor and premotor cortex: single-cell activity related to prior information about direction and extent of an intended movement. *J Neurophysiol* 61: 534–549, 1989.

468. Ringach DL, Hawken MJ, Shapley R. Dynamics of orientation tuning in macaque primary visual cortex. *Nature* 387: 281–284, 1997.

469. Roberts S. Isolation of an internal clock. *J Exp Psychol Anim Behav Proc* 7: 242–268, 1981.

470. Robbins TW, Arnsten AFT. The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. *Annu Rev Neurosci* 32: 267–287, 2009.

471. Robinson TE, Berridge KC. The neural basis for drug craving: an incentive-sensitization theory of addiction. *Brain Res Rev* 18: 247–291, 1993.

472. Robinson MJF, Berridge KC. Instant transformation of learned repulsion into motivational "wanting." *Curr Biol* 23: 282–289, 2013.

473. Rodriguez ML, Logue AW. Adjusting delay to reinforcement: comparing choice in pigeons and humans. *J Exp Psychol Anim Behav Proc* 14: 105–117, 1988.

474. Roelfsema PR, Tolboom M, Khayat PS. Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron* 56: 785–792, 2007.

475. Roesch MR, Calu DJ, Esber GR, Schoenbaum G. Neural correlates of variations in event processing during learning in basolateral amygdala. *J Neurosci* 30: 2464–2471, 2010.

476. Roesch MR, Olson CR. Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *J Neurophysiol* 90: 1766–1789, 2003.

477. Roesch MR, Olson CR. Neuronal activity dependent on anticipated and elapsed delay in macaque prefrontal cortex, frontal and supplementary eye fields, and premotor cortex. *J Neurophysiol* 94: 1469–1497, 2005.

478. Roesch MR, Olson CR. Neuronal activity in orbitofrontal cortex reflects the value of time. *J Neurophysiol* 94: 2457–2471, 2005.

479. Roesch MR, Singh T, Brown PL, Mullins SE, Schoenbaum G. Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. *J Neurosci* 29: 13365–13376, 2009.

480. Rogers QR, Harper AE. Selection of a solution containing histidine by rats fed a histidine-imbalanced diet. *J Comp Physiol Psychol* 72: 66–71, 1970.

481. Roiser JP, de Martino B, Tan GCY, Kumaran D, Seymour B, Wood NW, Dolan RJ. A genetically mediated bias in decision making driven by failure of amygdala control. *J Neurosci* 29: 5985–5991, 2009.

482. Roitman JD, Roitman MF. Risk-preference differentiates orbitofrontal cortex responses to freely chosen reward outcomes. *Eur J Neurosci* 31: 1492–1500, 2010.

483. Roitman JD, Shadlen MN. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22: 9475–9489, 2002.

484. Roitman MF, Wheeler RA, Carelli RM. Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. *Neuron* 45: 587–597, 2005.

485. Rolls ET, Critchley HD, Mason R, Wakeman EA. Orbitofrontal cortex neurons: role in olfactory and visual association learning. *J Neurophysiol* 75: 1970–1981, 1996.

486. Rolls ET, Grabenhorst F, Parris BA. Warm pleasant feelings in the brain. *NeuroImage* 41: 1504–1513, 2008.

487. Rolls ET, Yaxley S, Sienkiewicz ZJ. Gustatory responses of single neurons in the caudolateral orbitofrontal cortex of the macaque monkey. *J Neurophysiol* 64: 1055–1066, 1990.

488. Romo R, Brody CD, Hernández A, Lemus L. Neuronal correlates of parametricworking memory in the prefrontal cortex. *Nature* 399: 470–473, 1999.

489. Romo R, Hernández A, Zainos A. Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron* 41: 165–173, 2004.

490. Romo R, Hernández A, Zainos A, Lemus L, Brody CD. Neuronal correlates of decision-making in secondary somatosensory cortex. *Nat Neurosci* 5: 1217–1225, 2002.

491. Romo R, Schultz W. Neuronal activity preceding self-initiated or externally timed arm movements in area 6 of monkey cortex. *Exp Brain Res* 67: 656–662, 1987.

492. Romo R, Schultz W. Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *J Neurophysiol* 63: 592–606, 1990.

493. Romo R, Schultz W. Role of primate basal ganglia and frontal cortex in the internal generation of movements. III. Neuronal activity in the supplementary motor area. *Exp Brain Res* 91: 396–407, 1992.

494. Rorie AE, Gao J, McClelland JL, Newsome WT. Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS One* 5: e9308, 2010.

495. Rosenkranz JA, Grace AA. Dopamine-mediated modulation of odour-evoked amygdala potentials during pavlovian conditioning. *Nature* 417: 282–287, 2002.

496. Rossi MA, Fan D, Barter JW, Yin HH. Bidirectional modulation of substantia nigra activity by motivational state. *PLoS One* 8: e71598, 2013.

497. Rothschild M, Stiglitz IA. Increasing risk: definition. *J Econ Theory* 2: 225–243, 1970.

498. Rutledge RB, Skandalia N, Dayan P, Dolan RJ. A computational and neural model of momentary subjective well-being. *Proc Natl Acad Sci USA* 111: 12252–12257, 2014.

499. Salamone JD. The involvement of nucleus accumbens dopamine in appetitive and aversive motivation. *Behav Brain Res* 61: 117–133, 1994.

500. Samejima K, Ueda Y, Doya K, Kimura M. Representation of action-specific reward values in the striatum. *Science* 310: 1337–1340, 2005.

501. Samuelson P. A note on measurement of utility. *Rev Econ Stud* 4: 155–161, 1937.

502. Sato M, Hikosaka O. Role of primate substantia nigra pars reticulata in reward-oriented saccadic eye movement. *J Neurosci* 22: 2363–2373, 2002.

503. Satoh T, Nakai S, Sato T, Kimura M. Correlated coding of motivation and outcome of decision by dopamine neurons. *J Neurosci* 23: 9913–9923, 2003.

504. Savage LJ. *The Foundations of Statistics*. New York: Wiley, 1954.

505. Sawaguchi T, Goldman-Rakic PS. D1 dopamine receptors in prefrontal cortex: involvement in working memory. *Science* 251: 947–950, 1991.

506. Schall JD. Neuronal activity related to visually guided saccadic eye movements in the supplementary motor area of rhesus monkeys. *J Neurophysiol* 66: 530–558, 1991.

507. Schall JD, Stuphorn V, Brown JW. Monitoring and control of action by the frontal lobes. *Neuron* 36: 309–322, 2002.

508. Schall JD, Thompson KG. Neural selection and control of visually guided eye movements. *Annu Rev Neurosci* 22: 241–259, 1999.

509. Scherberger H, Andersen RA. Target selection signals for arm reaching in the posterior parietal cortex. *J Neurosci* 27: 2001–2012, 2007.

510. Schmalfuss B. The random attractor of the stochastic Lorenz system. *Z Angew Math Physik* 48: 951–975, 1997.

511. Schmidt R, Leventhal DK, Mallet N, Chen F, Berke JD. Canceling actions involves a race between basal ganglia pathways. *Nat Neurosci* 16: 1118–1124, 2013.

512. Schoenbaum G, Chiba AA, Gallagher M. Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *J Neurosci* 19: 1876–1884, 1999.

513. Schreiber CA, Kahneman D. Determinants of the remembered utility of aversive sounds. *J Exp Psych* 129: 27–42, 2000.

514. Schroeder T. *Three Faces of Desire*. Boston: MIT Press, 2004.

515. Schurger A, Jacobo Sitt JD D, Dehaene S. An accumulator model for spontaneous neural activity prior to self-initiated movement. *J Proc Natl Acad Sci USA* 109: E2904–E2913, 2012.

516. Schultz W. Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *J Neurophysiol* 56: 1439–1462, 1986.

517. Schultz W. Predictive reward signal of dopamine neurons. *J Neurophysiol* 80: 1–27, 1998.

518. Schultz W. Multiple dopamine functions at different time courses. *Annu Rev Neurosci* 30: 259–288, 2007.

519. Schultz W. Midbrain dopamine neurons: a retina of the reward system? In: *Neuroeconomics: Decision Making and the Brain*, edited by Glimcher PW, Camerer CF, Fehr E, Poldrack RA. New York: Academic, 2009, p. 323–329.

520. Schultz W. Potential vulnerabilities of neuronal reward, risk, and decision mechanisms to addictive drugs. *Neuron* 69: 603–617, 2011.

521. Schultz W, Apicella P, Ljungberg T. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J Neurosci* 13: 900–913, 1993.

522. Schultz W, Apicella P, Romo R, Scarnati E. Context-dependent activity in primate striatum reflecting past and future behavioral events. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, Beiser DG. Cambridge, MA: MIT Press, 1995, p. 11–28.

523. Schultz W, Apicella P, Scarnati E, Ljungberg T. Neuronal activity in monkey ventral striatum related to the expectation of reward. *J Neurosci* 12: 4595–4610, 1992.

524. Schultz W, Dayan P, Montague RR. A neural substrate of prediction and reward. *Science* 275: 1593–1599, 1997.

525. Schultz W, Romo R. Responses of nigrostriatal dopamine neurons to high intensity somatosensory stimulation in the anesthetized monkey. *J Neurophysiol* 57: 201–217, 1987.

526. Schultz W, Romo R. Neuronal activity in the monkey striatum during the initiation of movements. *Exp Brain Res* 71: 431–436, 1988.

527. Schultz W, Romo R. Dopamine neurons of the monkey midbrain: contingencies of responses to stimuli eliciting immediate behavioral reactions. *J Neurophysiol* 63: 607–624, 1990.

528. Schultz W, Romo R. Role of primate basal ganglia and frontal cortex in the internal generation of movements. I. Preparatory activity in the anterior striatum. *Exp Brain Res* 91: 363–384, 1992.

529. Searle JR. *Intentionality.* Cambridge, UK: Cambridge Univ. Press, 1983.

530. Seo H, Barraclough DJ, Lee D. Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cereb Cortex* 17: -i110–i117, 2007.

531. Seo H, Lee D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J Neurosci* 27: 8366–8377, 2007.

532. Seo H, Lee D. Behavioral and neural changes after gains and losses of conditioned reinforcers. *J Neurosci* 29: 3627–3641, 2009.

533. Seo M, Lee E, Averbeck BB. Action selection and action value in frontal-striatal circuits. *Neuron* 74: 947–960, 2012.

534. Seymour B, Daw ND, Roiser JP, Dayan P, Dolan R. Serotonin selectively modulates reward value in human decision-making. *J Neurosci* 32: 5833–5842, 2012.

535. Shadlen MN, Newsome WT. Neural basis of a perceptual decision in the parietal cortex (Area LIP) of the rhesus monkey. *J Neurophysiol* 86: 1916–1936, 2001.

536. Sheafor PJ. Pseudoconditioned jaw movements of the rabbit reflect associations conditioned to contextual background cues. *J Exp Psychol Anim Behav Proc* 104: 245–260, 1975.

537. Sheafor PJ, Gormezano I. Conditioning the rabbit's (*Oryctolagus cuniculus*) jaw-movement response: US magnitude effects on URs, CRs, and pseudo-CRs. *J Comp Physiol Psychol* 81: 449–456, 1972.

538. Shefrin HM, Thaler RH. The behavioral life-cycle hypothesis. *Econ Inq* 26: 609–643, 1988.

539. Sheffield FD, Roby TB. Reward value of a non-nutritive sweet taste. *J Comp Physiol Psychol* 43: 471–481, 1950.

540. Shen W, Flajolet M, Greengard P, Surmeier DJ. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* 321: 848–851, 2008.

541. Shidara M, Aigner TG, Richmond BJ. Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J Neurosci* 18: 2613–2625, 1998.

542. Shidara M, Richmond BJ. Anterior cingulate: single neuron signals related to degree of reward expectancy. *Science* 296: 1709–1711, 2002.

543. Shima K, Tanji J. Role for cingulate motor area cells in voluntary movement selection based on reward. *Science* 282: 1335–1338, 1998.

544. Simmons JM, Richmond BJ. Dynamic changes in representations of preceding and upcoming reward in monkey orbitofrontal cortex. *Cereb Cortex* 18: 93–103, 2008.

545. Simon HA. Rational choice and the structure of the environment. *Psychol Rev* 63: 129–138, 1956.

546. Singh S, Lewis RL, Barto AG. Where do rewards come from? In: *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, edited by Taatgen NAvan Rijn H. Austin, TX: Cogn. Sci. Soc., 2009, p. 2601–2606.

547. Singh S, Lewis RL, Barto AG, Sorg J. Intrinsically motivated reinforcement learning: an evolutionary perspective. *IEEE Trans Autonom Mental Dev* 2: 70–82, 2010.

548. Smith AD, Bolam JP. The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurones. *Trends Neurosci* 13: 259–265, 1990.

549. Smith PL, Ratcliff R. Psychology and neurobiology of simple decisions. *Trends Neurosci* 27: 161–168, 2004.

550. So NY, Stuphorn V. Supplementary eye field encodes option and action value for saccades with variable reward. *J Neurophysiol* 104: 2634–2653, 2010.

551. So NY, Stuphorn V. Supplementary eye field encodes reward prediction error. *J Neurosci* 32: 2950–2963, 2012.

552. Solomon RL, Corbit JD. An opponent-process theory of motivation. *Psychol Rev* 81: 119–145, 1974.

553. Solomon PR, Vander Schaaf ER, Thompson RF, Weisz DJ. Hippocampus and trace conditioning of the rabbit's classically conditioned nictitating membrane response. *Behav Neurosci* 100: 729–744, 1986.

554. Soltani A, Lee D, Wang XJ. Neural mechanism for stochastic behaviour during a competitive game. *Neur Networks* 19: 1075–1090, 2006.

555. Soon CS, Brass M, Heinze HJ, Haynes JD. Unconscious determinants of free decisions in the human brain. *Nat Neurosci* 11: 543–545, 2008.

556. Stalnaker TA, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G. Reward prediction error signaling in posterior dorsomedial striatum is action specific. *J Neurosci* 32: 10296–10305, 2012.

557. Stanisor L, van der Togt C, Cyriel MA, Pennartz CMA, Roelfsema PR. A unified selection signal for attention and reward in primary visual cortex. *Proc Natl Acad Sci USA* 110: 9136–9141, 2013.

558. Stanton SJ, Mullette-Gillman O'DA, McLaurin RE, Kuhn CM, LaBar KS, Platt ML, Huettel SA. Low- and high-testosterone individuals exhibit decreased aversion to economic risk. *Psychol Sci* 22: 447–453, 2011.

559. Stauffer WR, Lak A, Bossaerts P, Schultz W. Economic choices reveal probability distortion in monkeys. *J Neurosci* 35: 3146–3154, 2015.

560. Stauffer WR, Lak A, Schultz W. Dopamine reward prediction error responses reflect marginal utility. *Curr Biol* 24: 2491–2500, 2014.

561. Stein RB. Some models of neuronal variability. *Biophys J* 7: 37–68, 1967.

562. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* 16: 966–973, 2013.

563. Steinfels GF, Heym J, Strecker RE, Jacobs BL. Behavioral correlates of dopaminergic unit activity in freely moving cats. *Brain Res* 258: 217–228, 1983.

564. Stevens CF. Inferences about membrane properties from electrical noise measurements. *Biophys J* 12: 1028–1047, 1972.

565. Stone M. Models for choice reaction time. *Psychometrika* 25: 251–260, 1960.

566. Stopper CM, Tse MTL, David Montes DR R, Wiedman CR, Floresco SB. Overriding phasic dopamine signals redirects action selection during risk/reward decision making. *Neuron* 84: 177–189, 2014.

567. Strassberg AF, DeFelice LJ. Limitations of the Hodgkin-Huxley formalism: effects of single channel kinetics on transmembrane voltage dynamics. *Neur Comp* 5: 843–855, 1993.

568. Stuber GD, Klanker M, de Ridder B, Bowers MS, Joosten RN, Feenstra MG, Bonci A. Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. *Science* 321: 1690–1692, 2008.

569. Stuphorn V, Taylor TL, Schall JD. Performance monitoring by the supplementary eye field. *Nature* 408: 857–860, 2000.

570. Sugam JA, Day JJ, Wightman RM, Carelli RM. Phasic nucleus accumbens dopamine encodes risk-based decision-making behavior. *Biol Psychiat* 71: 199–215, 2012.

571. Sugrue LP, Corrado GS, Newsome WT. Matching behavior and the representation of value in the parietal cortex. *Science* 304: 1782–1787, 2004.

572. Sul JH, Kim H, Huh N, Lee D, Jung MW. Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron* 66: 449–460, 2010.

573. Suri R, Schultz W. A neural network with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience* 91: 871–890, 1999.

574. Sutton RS, Barto AG. Toward a modern theory of adaptive networks: expectation and prediction. *Psychol Rev* 88: 135–170, 1981.

575. Sutton RS, Barto AG. *Reinforcement Learning*. Cambridge, MA: MIT Press, 1998.

576. Tai LH, Lee AM, Benavidez N, Bonci A, Wilbrecht L. Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat Neurosci* 15: 1281–1289, 2012.

577. Takahashi T. A mathematical framework for probabilistic choice based on information theory and psychophysics. *Med Hypoth* 67: 183–186, 2006.

578. Takahashi YK, Roesch MR, Stalnaker TA, Haney RZ, Calu DJ, Taylor AR, Burke KA, Schoenbaum G. The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. *Neuron* 62: 269–280, 2009.

579. Takikawa Y, Kawagoe R, Hikosaka O. A possible role of midbrain dopamine neurons in short- and long-term adaptation of saccades to position-reward maping. *J Neurophysiol* 92: 2520–2529, 2004.

580. Talwar SK, Xu S, Hawley ES, Weiss SA, Moxon KA, Chapin JK. Rat navigation guided by remote control. *Nature* 417: 37–38, 2002.

581. Tan CO, Bullock D. A local circuit model of learned striatal and dopamine cell responses under probabilistic schedules of reward. *J Neurosci* 28: 10062–10074, 2008.

582. Tan KR, Yvon C, Turiault M, Mirzabekov JJ, Doehner J, Labouèbe G, Deisseroth K, Tye KM, Lüscher C. GABA neurons of the VTA drive conditioned place aversion. *Neuron* 73: 1173–1183, 2012.

583. Tang KC, Low MJ, Grandy DK, Lovinger DM. Dopamine-dependent synaptic plasticity in striatum during in vivo development. *Proc Natl Acad Sci USA* 98: 1255–1260, 2001.

584. Tanimoto H, Heisenberg M, Gerber B. Event timing turns punishment to reward. *Nature* 430: 983, 2004.

585. Teodorescu TR, Usher M. Disentangling decision models: from independence to competition. *Psychol Rev* 120: 1–38, 2013.

586. Tesauro G. TD-Gammon, a self-teaching backgammon program, achieves master-level play. *Neural Comp* 6: 215–219, 1994.

587. Thompson KG, Bichot NP, Schall JD. Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *J Neurophysiol* 77:1046–1050, 1997.

588. Thompson KG, Hanes DP, Bichot NP, Schall JD. Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *J Neurophysiol* 76: 4040–4055, 1996.

589. Thorndike EL. *Animal Intelligence: Experimental Studies*. New York: MacMillan, 1911.

590. Thut G, Schultz W, Roelcke U, Nienhusmeier M, Maguire RP, Leenders KL. Activation of the human brain by monetary reward. *NeuroReport* 8: 1225–1228, 1997.

591. Tindell AJ, Smith KS, Peciña S, Berridge KC, Aldridge JW. Ventral pallidum firing codes hedonic reward: when a bad taste turns good. *J Neurophysiol* 96: 2399–2409, 2006.

592. Thorpe SJ, Rolls ET, Maddison S. The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp Brain Res* 49: 93–115, 1983.

593. Tinklepaugh OL. An experimental study of representation factors in monkeys. *J Comp Psychol* 8: 197–236, 1928.

594. Toan DL, Schultz W. Responses of rat pallidum cells to cortex stimulation and effects of altered dopaminergic activity. *Neuroscience* 15: 683–694, 1985.

595. Tobler PN, Christopoulos GI, O'Doherty JP, Dolan RJ, Schultz W. Neuronal distortions of reward probability without choice. *J Neurosci* 28:11703–11711, 2008.

596. Tobler PN, Christopoulos GI, O'Doherty JP, Dolan RJ, Schultz W. Risk-dependent reward value signal in human prefrontal cortex. *Proc Natl Acad Sci USA* 106: 7185–7190, 2009.

597. Tobler PN, Dickinson A, Schultz W. Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *J Neurosci* 23: 10402–10410, 2003.

598. Tobler PN, Fiorillo CD, Schultz W. Adaptive coding of reward value by dopamine neurons. *Science* 307: 1642–1645, 2005.

599. Tobler PN, O'Doherty JP, Dolan R, Schultz W. Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J Neurophysiol* 97: 1621–1632, 2007.

600. Tremblay L, Hollerman JR, Schultz W. Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J Neurophysiol* 80: 964–977, 1998.

601. Tremblay L, Schultz W. Relative reward preference in primate orbitofrontal cortex. *Nature* 398: 704–708, 1999.

602. Tremblay L, Schultz W. Reward-related neuronal activity during go-nogo task performance in primate orbitofrontal cortex. *J Neurophysiol* 83: 1864–1876, 2000.

603. Tremblay L, Schultz W. Modifications of reward expectation-related neuronal activity during learning in primate orbitofrontal cortex. *J Neurophysiol* 83: 1877–1885, 2000.

604. Tsai CT, Nakamura S, Iwama K. Inhibition of neuronal activity of the substantia nigra by noxious stimuli and its modification by the caudate nucleus. *Brain Res* 195: 299–311, 1980.

605. Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K. Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science* 324: 1080–1084, 2009.

606. Tsujimoto S, Genovesio A, Wise SP. Monkey orbitofrontal cortex encodes response choices near feedback time. *J Neurosci* 29: 2569–2574, 2009.

607. Umeno MM, Goldberg ME. Spatial processing in the monkey frontal eye field. I. Predictive visual responses. *J Neurophysiol* 78: 1373–1383, 1997.

608. Ungless MA, Magill PJ, Bolam JP. Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science* 303: 2040–2042, 2004.

609. Usher M, McClelland JL. The time course of perceptual choice: the leaky, competing accumulator model. *Psych Rev* 108: 550–592, 2011.

610. Valenza E, Simion F, Macchi Cassia V, Umiltà C. Face Preference at Birth. *J Exp Psych Hum Percept Perform* 22: 892–903, 1996.

611. Van Vreeswijk C, Sompolinsky H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1726, 1996.

612. Van Wolkenten M, Brosnan SF, de Waal F. Inequity responses of monkeys modified by effort. *Proc Nat Acad Sci USA* 104: 18854–18859, 2007.

613. Van Zessen R, Phillips JL, Budygin EA, Stuber GD. Activation of VTA GABA neurons disrupts reward consumption. *Neuron* 73: 1184–1194, 2012.

614. Vickers D. Evidence for an accumulator model of psychophysical discrimination. *Ergonomics* 13: 37–58, 1970.

615. Vickery TJ, Chun MM, Lee D. Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron* 72: 166–177, 2011.

616. Vijayraghavan S, Wang M, Birnbaum SG, Williams GV, Arnsten AFT. Inverted-U dopamine D1 receptor actions on prefrontal neurons engaged in working memory. *Nat Neurosci* 10: 376–384, 2007.

617. Von Neumann J, Morgenstern O. *The Theory of Games and Economic Behavior*. Princeton, NJ: Princeton Univ. Press, 1944.

618. Waelti P, Dickinson A, Schultz W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412: 43–48, 2001.

619. Wald A, Wolfowitz J. Optimum character of the sequential probability ratio test. *Ann Math Statist* 19: 326–339, 1947.

620. Wallis JD, Miller EK. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur J Neurosci* 18: 2069–2081, 2003.

621. Wanat MJ, Kuhnen CM, Phillips PEM. Delays conferred by escalating costs modulate dopamine release to rewards but not their predictors. *J Neurosci* 30: 12020–12027, 2010.

622. Wang XJ. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36: 955–968, 2002.

623. Wang XJ. Decision making in recurrent neuronal circuits. *Neuron* 60: 215–234, 2008.

624. Wang LP, Li F, Wang D, Xie K, Wang D, Shen X, Tsien JZ. NMDA receptors in dopaminergic neurons are crucial for habit learning. *Neuron* 72: 1055–1066, 2011.

625. Watabe-Uchida M, Zhu L, Ogawa SK, Vamanrao A, Uchida N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74, 858–873, 2012.

626. Watanabe M. Prefrontal unit activity during associative learning in the monkey. *Exp Brain Res* 80: 296–309, 1990.

627. Watanabe M. Reward expectancy in primate prefrontal neurons. *Nature* 382: 629–632, 1996.

628. Watanabe M, Hikosaka K, Sakagami M, Shirakawa SI. Coding and monitoring of behavioral context in the primate prefrontal cortex. *J Neurosci* 22: 2391–2400, 2002.

629. Watson KK, Platt ML. Social signals in primate orbitofrontal cortex. *Curr Biol* 22: 2268–2273, 2012.

630. Weber BJ, Chapman GB. Playing for peanuts: why is risk seeking more common for low-stakes gambles? *Organiz Behav Human Dec Proc* 97: 31–46, 2005.

631. Weber EU, Johnson EJ. Decisions under uncertainty: psychological, economic, and neuroeconomic explanations of risk preference. In: *Neuroeconomics* , edited by Glimcher PW, Camerer CF, Fehr E, Poldrack RA. London: Academic, 2009.

632. Weber EU, Milliman RA. Perceived risk attitudes: relating risk perception to risky choice. *Management Sci* 43: 123–144, 1997.

633. Weber EU, Shafir S, Blais AR. Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation. *Psychol Rev* 111: 430–445, 2004.

634. Wegner DM. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press, 2002.

635. West EA, Forcelli PA, McCuea DL, Malkova L. Differential effects of serotonin-specific and excitotoxic lesions of OFC on conditioned reinforcer devaluation and extinction in rats. *Behav Brain Res* 246: 10–14, 2013.

636. Weinrich M, Wise SP. The premotor cortex of the monkey. *J Neurosci* 2: 1329–1345, 1982.

637. Wickens J, Kötter R. Cellular models of reinforcement. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, Beiser DG. Cambridge, MA: MIT Press, 1995, p. 187–214.

638. Wightman RM, Robinson DL. Transient changes in mesolimbic dopamine and their association with "reward." *J Neurochem* 82: 721–735, 2002.

639. Williams ZM, Eskandar EN. Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nat Neurosci* 4: 562–568, 2006.

640. Wise RA. Brain reward circuitry: insights from unsensed incentives. *Neuron* 36: 229–240, 2002.

641. Witten IB, Steinberg EE, Lee SY, Davidson TJ, Zalocusky KA, Brodsky M, Yizhar O, Cho SL, Gong S, Ramakrishnan C, Stuber GD, Tye KM, Janak PH, Deisseroth K. Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron* 72: 721–733, 2011.

642. Yagishita S, Hayashi-Takagi A, Ellis-Davies GCR, Urakubo H, Ishii S, Kasai H. A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* 345: 1616–1620, 2014.

643. Yang T, Shadlen MN. Probabilistic reasoning by neurons. *Nature* 447: 1075–1080, 2007.

644. Yamada H, Tymula A, Louie K, Glimcher PW. Thirst-dependent risk preferences in monkeys identify a primitive form of wealth. *Proc Natl Acad Sci USA* 110: 15788–15793, 2013.

645. Yasuda M, Yamamoto S, Hikosaka O. Robust representation of stable object values in the oculomotor basal ganglia. *J Neurosci* 32: 16917–16932, 2012.

646. Yin HH, Ostlund SB, Knowlton BJ, Balleine BB. The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci* 23: 513–523, 2005.

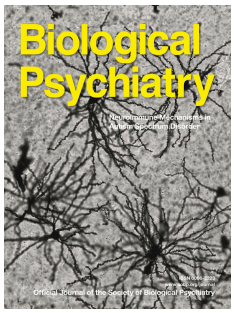
647. Yoshida K, Saito N, Iriki A, Isoda M. Social error monitoring in macaque frontal cortex. *Nat Neurosci* 15, 1307–1312, 2012.

648. Zhang JC, Lau PM, Bi GQ. Gain in sensitivity and loss in temporal contrast of STDP by dopaminergic modulation at hippocampal synapses. *Proc Natl Acad Sci USA* 106: 1328–1333, 2009.

649. Zweifel LS, Argilli E, Bonci A, Palmiter R. Role of NMDA receptors in dopamine neurons for plasticity and addictive behaviors. *Neuron* 59: 486–496, 2008.

650. Zweifel LS, Parker JG, Lobb CJ, Rainwater A, Wall VZ, Fadok JP, Darvas M, Kim MJ, Mizumori SJ, Paladini CA, Philipps PEM, Palmiter R. Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. *Proc Natl Acad Sci USA* 106, 7281–7288, 2009.

Accepted Manuscript



Neurobiological markers of resilience to depression and anxiety following childhood maltreatment: The role of neural circuits supporting the cognitive control of emotion

Alexandra M. Rodman, Jessica L. Jenness, David G. Weissman, Daniel S. Pine, Katie A. McLaughlin

PII: S0006-3223(19)31333-2

DOI: https://doi.org/10.1016/j.biopsych.2019.04.033

Reference: BPS 13854

To appear in: *Biological Psychiatry*

Received Date: 15 January 2019

Revised Date: 11 April 2019

Accepted Date: 15 April 2019

Please cite this article as: Rodman A.M., Jenness J.L., Weissman D.G., Pine D.S. & McLaughlin K.A., Neurobiological markers of resilience to depression and anxiety following childhood maltreatment: The role of neural circuits supporting the cognitive control of emotion, *Biological Psychiatry* (2019), doi: https://doi.org/10.1016/j.biopsych.2019.04.033.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Special issue: Developmental Neurobiological Perspectives on Resilience and Risk for Anxiety and Depression in *Biological Psychiatry.*

Neurobiological markers of resilience to depression and anxiety following childhood maltreatment: The role of neural circuits supporting the cognitive control of emotion

Alexandra M. Rodman[1], Jessica L. Jenness[2], David G. Weissman[1], Daniel S. Pine[3] & Katie A. McLaughlin[1]

Author Affiliations:

1. Department of Psychology, Harvard University

2. Department of Psychiatry and Behavioral Sciences, University of Washington

3. The National Institute for Mental Health, USA

Address correspondence to:

Alexandra M. Rodman
Department of Psychology
Harvard University
William James Hall Room 1002
33 Kirkland Street
Cambridge MA 02138

arodman@fas.harvard.edu

Keywords: Resilience, Neurobiological, Adversity, Emotion Regulation, Depression, Adolescence
Abstract Word Count: 245
Main Text Word Count: 3,605 (excluding references, captions and supplement)
Figures: 4 (+1 Supplement)
Tables: 1 (+3 Supplement)
Supplement: 1
References: 121

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

Childhood adversity is strongly linked to negative mental health outcomes, including depression and anxiety. Leveraging cognitive neuroscience to identify mechanisms that contribute to resilience in children with a history of maltreatment may provide viable intervention targets for the treatment or prevention of psychopathology. We present a conceptual model of a potential neurobiological mechanism of resilience to depression and anxiety following childhood adversity. Specifically, we argue that neural circuits underlying the cognitive control of emotion may promote resilience, wherein a child's ability to recruit the frontoparietal control network to modulate amygdala reactivity to negative emotional cues—such as during cognitive reappraisal— buffers risk for internalizing symptoms following exposure to adversity. We provide preliminary support for this model of resilience in a longitudinal sample of 151 participants aged 8-17 years with (n=79) and without (n=72) a history of childhood maltreatment who completed a cognitive reappraisal task while undergoing fMRI. Among maltreated youth, those who were better able to recruit prefrontal control regions and modulate amygdala reactivity during reappraisal exhibited lower risk for depression over time. By contrast, no association was observed between neural functioning during reappraisal and depression among youth without a history of maltreatment. These preliminary findings support the hypothesis that children who are better able to regulate emotion through recruitment of the frontoparietal network exhibit greater resilience to depression following childhood maltreatment. Interventions targeting cognitive reappraisal and other cognitive emotion regulation strategies may have potential for reducing vulnerability to depression among children exposed to adversity.

Childhood adversity, which robustly predicts psychopathology (1–3), refers to negative experiences that deviate from the expectable environment, requiring meaningful adaptation by an average child (4). These experiences can reflect either *threat*, defined as relating to harmful experiences, or *deprivation*, the absence of expected environmental inputs, such as caregiver support and cognitive stimulation (5–7). The current paper focuses on childhood adversity in the form of threat, specifically maltreatment (e.g., physical or sexual abuse), as it has particularly strong ties to depression and anxiety (3, 8–10).

Although childhood adversity is a powerful predictor of psychopathology, this relationship is not deterministic; many children who have encountered severe forms of adversity demonstrate resilience and do not go on to develop mental health problems (11–15). Resilience involves processes that buffer children from risk for these negative consequences (16). Identifying mechanisms of resilience may reveal targets for preventative interventions designed to protect children following adversity (4). Considerable work examines factors that promote resilience (15, 16), but few studies examine neurobiological mechanisms conferring resilience among children exposed to threat-related adversity. Here, we advance a neurobiological model of resilience, focusing on neural circuits underlying the cognitive control of emotion. Specifically, we posit that a child's ability to recruit the frontoparietal control network to modulate amygdala reactivity to negative emotional cues—such as during cognitive reappraisal and other effortful forms of emotion regulation— buffers risk for internalizing symptoms following exposure to adversity. We provide preliminary data testing this proposed framework using neuroimaging data of cognitive reappraisal in a longitudinal sample of children and adolescents exposed to maltreatment.

Defining Resilience

Developmental and clinical psychologists have long been interested in resilience, and various definitions have been proposed. Some have conceived of resilience as a fixed trait or set of traits that are immutable (17–19) and may be present within an individual whether or not they have experienced adversity (20). Instead, we utilize the definition that resilience reflects an absence of negative outcomes despite exposure to adversity (15, 16, 21). To study this form of resilience from an empirical perspective, one must identify specific factors or developmental processes that moderate the association between adversity and negative outcomes, such that the adversity-negative outcome relationship is weaker among those who have higher levels of the resilience factor.

Resiliency factors can occur at multiple interacting levels of the bioecological milieu (22–24), from entire cultures (25) to neighborhoods (26) to families (27) to children's temperaments (28) all the way down to individual genes (29). In addition, the factors that lead to resilience may depend on the nature of the adversity experienced and the social and cultural context to which an individual must adapt (24). Developmental cognitive neuroscience may generate unique insights into resiliency factors. Work in this area can be leveraged to identify resilience mechanisms at the level of specific cognitive-affective processes and their underlying neural networks that confer protection against psychopathology following experiences of adversity. Moreover, a rich history in this area informs research performed at the therapeutic level (30). Identifying specific neurobiological, cognitive, and affective mechanisms of resilience that are modifiable is critically important for informing models of risk and resilience, as well as identifying viable targets for intervention to treat or prevent psychopathology (4).

Here, we advance the possibility that the effective engagement of cognitive control networks in service of modulating negative emotions may be a neurobiological mechanism of resilience to depression and anxiety following childhood adversity. Neural circuits underlying the cognitive control of emotion, including the frontoparietal control network, have been most studied in relation to the emotion regulation strategy of cognitive reappraisal.

Cognitive Reappraisal

Cognitive reappraisal involves thinking about a stimulus in a way that changes the meaning to modify one's emotional response (e.g., to reduce negative or enhance positive emotion) (31–36). Cognitive reappraisal has been shown to modulate emotional responses in experimental settings (31, 32, 37–40), real-world settings (38, 41), and clinical intervention studies, where training to enhance reappraisal is associated with reductions in symptoms of depression and anxiety in children and adults (42–44).

Functional magnetic resonance imaging (fMRI) studies reveal a network of brain regions recruited during cognitive reappraisal that modulate amygdala activation (45). These studies typically use specific reappraisal strategies, such as psychological distancing or reinterpretation, as tactics for reducing emotional responses to negative stimuli (e.g., images of a car crash) (31, 34, 35, 46–49). Meta-analysis shows that during cognitive reappraisal compared to passive viewing of emotional stimuli, regulatory regions of the frontoparietal network are engaged and modulate amygdala activity (50). The frontoparietal regions recruited during cognitive reappraisal are broadly involved in cognitive control (51, 52) and include multiple prefrontal regions in both dorsal and ventral areas, the dorsal anterior cingulate, as well as posterior association cortex encompassing inferior parietal sulcus (31, 34, 50, 53–57). Recruitment of the

frontoparietal network during reappraisal may serve to select and maintain reappraisal-related features and goals while engaging in and monitoring progress of the construction of a new appraisal (45, 56).

Studies examining reappraisal in youth have shown that children as young as six years old can successfully employ this technique (57). Moreover, the extent of reappraisal success appears to improve linearly with age (36, 57–59) in association with increasing recruitment of prefrontal regions and diminishing activation in the amygdala, as well as greater inverse coupling between prefrontal regions and the amygdala (57). It should be noted, however, that some studies show behavioral indices of reappraisal success appear similar across development (60, 61), which could be explained by the type of reappraisal tactic being used (e.g., reinterpretation vs. distancing), as reinterpretation requires more complex, higher-order thinking.

Both adults and youth express patterns of frontoparietal recruitment during reappraisal in ways that modulate amygdala activation (36, 57, 59, 62–64). Given the amygdala's role in the processing of salient events (65–67), levels of amygdala modulation may reflect successful regulation of affective responding, although the specific patterns of connectivity that underlie this modulation remain a source of debate. Because lateral prefrontal regions have sparse direct projections to the amygdala, some studies suggest that activation in these regions modulates amygdala function via projections through the more densely-connected medial prefrontal cortex (53, 57, 68–72). Alternatively, lateral prefrontal regions may modulate amygdala activity via projections through lateral temporal cortex regions involved in semantic representation (34, 39, 50, 64). In either case, stronger inverse coupling between the prefrontal cortex and the amygdala is believed to produce greater reductions in negative emotion during reappraisal (53, 55). Below, we argue that the ability to successfully modulate the amygdala by recruiting this cognitive

control circuitry in the service of reappraisal is a key neurobiological mechanism of resilience to depression and anxiety following experiences of adversity.

Cognitive Reappraisal Neural Circuitry in Depression and Anxiety

Behaviorally, children and adults with depression and anxiety report similar reductions in negative emotion following reappraisal as those without psychopathology (31, 48, 73–77). However, those with depression and anxiety appear to use less efficient reappraisal strategies (73, 74). Disruptions in neural activation of frontoparietal and limbic regions involved in cognitive reappraisal have also been associated with depression and anxiety. However, findings vary across age and diagnosis. Some studies find affected relative to unaffected individuals to show greater recruitment of frontoparietal regions and heightened amygdala activity during cognitive reappraisal (48, 69, 73, 76), whereas others find affected individuals to show reduced recruitment across prefrontal regions (77, 78) or no amygdala differences (31, 75, 78). Taken together, data suggest that affected individuals manifest some form of disrupted prefrontal capacity to modulate the amygdala, which may reflect less efficient recruitment of regions supporting reappraisal processes.

Cognitive Control Circuitry as a Mechanism of Resilience following Childhood Adversity

The ability to effectively recruit frontoparietal circuitry in support of effortful emotion regulation strategies, such as cognitive reappraisal, could be a critical compensatory mechanism that may help to buffer against the heightened emotional and neurobiological reactivity that has been commonly observed following maltreatment and other forms of childhood adversity. Prior work consistently demonstrates that children exposed to adversity, particularly experiences of

threat, exhibit elevated emotional responses to negative stimuli assessed at multiple levels of analysis, including subjective report (79–82), autonomic nervous system response (83, 84), and amygdala reactivity (49, 85–89), the latter of which has also been confirmed in meta-analysis (90). Heightened emotional reactivity is a well-established risk factor for the emergence of depression and anxiety in youth (80, 81, 91–96). The ability to recruit frontoparietal circuitry to modulate amygdala reactivity may buffer children from internalizing problems that arise following these adversity-related increases in emotional reactivity. Indeed, among children exposed to adversity, preliminary evidence links greater structural and functional integrity within emotion regulatory circuits—encompassing connections between multiple regions of the medial prefrontal cortex and the amygdala—to resilience, in the form of lower risk for negative mental and physical health outcomes (97–100). These studies provide preliminary support for the notion that the ability to recruit prefrontal circuitry to modulate amygdala responses may be a neurobiological mechanism underlying adaptation to childhood adversity that dampens emotional and neurobiological hyper-reactivity to salient environmental stimuli, ultimately contributing to lower risk for psychopathology.

Effective cognitive reappraisal capacity may also be particularly important for children who have experienced adversity due to the strong continuity between exposure to adversity and subsequent exposure to stressful life events. Experiences of childhood adversity are highly co-occurring, such that children who experience one form of adversity (e.g., sexual abuse) typically experience several others (e.g., neglect and domestic violence) (1, 3). In addition, children exposed to adversity experience higher levels of stressful life events and chronic stress across academic, peer, and family domains (98, 101). Exposure to stressful life events and chronic stressors are well-established risk factors for depression and anxiety (102–105), and the link

between stressful life events and internalizing psychopathology is stronger among those who have experienced childhood adversity (101, 106–109). The ability to effectively utilize cognitive reappraisal may protect children from the negative mental health consequences of ongoing exposure to stressors, as has been shown in adults (110). Indeed, animal models of stress in non-human primate research support the notion that enhanced recruitment of regulatory circuits may promote resilience following early life stress (e.g., maternal separation; 111). For children who have experienced maltreatment, the capacity to flexibly deploy cognitive reappraisal and underlying control circuitry may be all the more important, considering the high likelihood that these children live in chronically stressful environments that frequently elicit negative emotions. However, no study to date has explicitly tested whether children's ability to explicitly engage effortful strategies and recruit prefrontal circuitry to modulate amygdala responses to negative emotional cues moderates risk for depression and anxiety following adversity.

Evaluating a Neurobiological Mechanism of Resilience

Using preliminary data from our group, we completed a set of exploratory analyses to provide an empirical test of our proposed mechanism of resilience. Specifically, we assessed whether the ability to modulate amygdala reactivity using cognitive reappraisal is a potential neurobiological mechanism of resilience to depression and anxiety among children exposed to maltreatment, a form of adversity that has particularly strong associations with internalizing psychopathology. If the proposed neurobiological mechanism of resilience is valid, we should expect that the association between child maltreatment and depression and anxiety symptoms will be lower among children and adolescents who exhibit: 1) greater modulation of amygdala responses to negative stimuli using cognitive reappraisal; 2) greater recruitment of prefrontal

regions known to be engaged during successful cognitive reappraisal; and 3) a greater tendency to use reappraisal strategies in their daily lives.

We examined these hypotheses in a longitudinal sample of 151 participants aged 8-17 with (n=79) and without (n=72) history of childhood maltreatment (e.g., physical or sexual abuse) who completed an emotion regulation task while undergoing fMRI. This task assessed neural activation during passive viewing and effortful attempts to regulate emotional responses to negative stimuli using cognitive reappraisal (Figure 1). Participants also reported on their tendency to engage in reappraisal in their daily lives. Symptoms of depression and anxiety were assessed at the time of the initial neuroimaging assessment and at a follow up assessment approximately two years later. See Table 1 for participant characteristics. All analyses were completed controlling for sex, age, race/ethnicity, and socioeconomic status (the income-to-needs ratio). We found no evidence for these patterns of resilience in relation to symptoms of anxiety. Below, we report the pattern of results for depression symptoms. For more details on participants, inclusion and exclusion criteria of the study, task design, measures, and analytical approach, see *Supplemental Materials.*

** TABLE 1 **

** FIGURE 1 **

Preliminary Evidence

Across the entire sample, the use of cognitive reappraisal elicited the expected pattern of activation of frontoparietal regions and decreased activation of the amygdala (Supplemental Figure 1). Our primary hypothesis was that maltreated children who exhibited greater

modulation of the amygdala during reappraisal (i.e., lower amygdala activity during reappraisal relative to passive viewing of negative stimuli) would be at lower risk for developing depression. To evaluate this hypothesis, we examined whether reappraisal-related amygdala modulation moderated the association between child maltreatment and depression symptoms and observed a significant maltreatment-by-brain function interaction (b=4.20, t=2.41, p=.018). Greater reappraisal-related amygdala modulation predicted lower risk for depression at follow up among maltreated youth but had no relationship to depression among those without a history of maltreatment (Figure 2), suggesting that the ability to modulate amygdala responses using cognitive reappraisal may be a marker of resilience. This finding remained significant after numerous sensitivity analyses (e.g., including pubertal stage as a covariate rather than age, including IQ as a covariate) and survived correction for multiple comparisons. Greater information on the analytical approach and sensitivity analyses can be found in *Supplemental Materials.*

** FIGURE 2 **

To follow up this finding, we conducted a set of exploratory analyses examining whether reappraisal-related recruitment of prefrontal regions moderated the association between child maltreatment and depression symptoms. To do so, we examined a set of prefrontal regions that were engaged during reappraisal relative to passive viewing of emotional stimuli in the whole sample (Supplemental Figure 1). This analysis revealed a significant maltreatment-by-brain function interaction in two regions consistently implicated in cognitive reappraisal (50): the right superior frontal gyrus (SFG by maltreatment interaction: b=-3.11, t=-2.50, p=.014) and right dorsal anterior cingulate cortex (dACC by maltreatment interaction: b=-2.65, t=-1.95, p=.052).

In both cases, greater recruitment during reappraisal predicted lower depression symptoms at baseline, but only for those with a history of maltreatment (Figure 3). Due to the exploratory nature of these analyses, the results presented here were not corrected for multiple comparisons. However, these patterns persist in multiple sensitivity analyses examining additional covariates (see *Supplemental Materials*).

FIGURE 3

Finally, we assessed of whether greater use of cognitive reappraisal as an emotion regulation strategy in everyday life moderated the association between child maltreatment and depression symptoms. Again, we observed a significant maltreatment-by-use interaction in the expected direction (b=-0.044, t=-2.420, p=.017), whereby the association between severity of child maltreatment and depression symptoms was weaker among those who reported greater use of cognitive reappraisal (Figure 4). This finding suggests that not only is the efficacy of cognitive reappraisal associated with resilience to depression following child maltreatment, but also the tendency to use reappraisal techniques in everyday life.

** FIGURE 4 **

Discussion

The ability to modulate negative emotion using cognitive control strategies may represent a resiliency marker, which protects against depression in children who have experienced adversity. The current report finds evidence of such a relationship. Specifically, greater capacity to modulate amygdala activation using cognitive reappraisal predicts decreasing levels of depressive symptoms across a two-year follow-up period. Similarly, greater recruitment of

prefrontal regions also predicts lower concurrent symptoms of depression among children with history of maltreatment. Finally, maltreated children who reported a greater tendency to use reappraisal as a coping strategy in everyday life have lower levels of depressive symptoms than those who did not. Of note, in youth unexposed to maltreatment, neither amygdala modulation, prefrontal function, nor reported use of cognitive reappraisal in daily life related to symptoms of depression. These preliminary findings support the proposed model of resilience. This model underscores specific neurobiological factors involved in the cognitive control of emotion as a potential protective factor buffering children who have experienced adversity from negative outcomes later in life.

Examining these questions in a youth sample is particularly advantageous, as insights about mechanisms of resilience can be leveraged to inform early interventions. Additionally, studies examining resilience to psychopathology in adulthood following childhood adversity often reflect differences in the accumulation of environmental stressors over the life course and suffer from recall biases that are mitigated, at least somewhat, when studying resilience in closer proximity to the initial source of adversity. The major contribution from the current study relates to the contrast of brain-behavior associations in youth with and without maltreatment. Specific neural markers of effective cognitive reappraisal only related to depressive symptoms in children with a history of maltreatment. As such, the current study delineated a marker of resilience.

Resilience involves many dynamic and interacting factors that modulate risk in the face of adversity (24), including cultural (25), familial (27), and genetic factors (29). The current study leverages advances in developmental cognitive neuroscience to examine brain function supporting specific cognitive processes as a mechanistic path to resilience. An analysis of resilience on this neural-psychological level is advantageous as it can be used to identify

malleable targets for preventing the onset or progression of internalizing disorders. This approach has been successfully undertaken in the context of anxiety and posttraumatic stress disorders, wherein basic mechanistic understanding of threat-related biases in information processing and associated neural functioning has shaped therapeutic approaches (e.g., attention bias modification therapy, ABMT) aimed at mitigating anxiety and trauma-related symptoms (30, 112–115). ABMT was designed to explicitly target the types of attentional biases towards threat that characterize anxiety disorders by training subjects to orient attention away from threatening cues in the task (116, 117). As such, it is a model of how basic understanding of neurobiological mechanisms can directly inform preventative and treatment approaches (30, 114, 118).

Similar translational approaches leveraging neurocognitive understanding of cognitive reappraisal could be useful in preventative and intervention efforts aimed at mitigating risk for internalizing disorders, especially in children who have experienced adversity. Cognitive reappraisal is a core intervention technique central to evidence-based psychotherapy practices for depression and anxiety, such as cognitive behavioral therapy (CBT) (42). Moreover, CBT for depression and anxiety has been associated with changes in frontoamygdala neural circuitry, including reductions in amygdala hyperresponsiveness and increased engagement of medial and lateral regions of the prefrontal cortex during both resting state and task performance (119, 120). Further, our findings show that in addition to cognitive reappraisal ability, the tendency to use it in everyday life may also serve as an important buffer for depression following adversity. Therefore, clinicians may incorporate training geared towards scaffolding and encouraging the use of reappraisal in the daily lives of children exposed to adversity. These strategies may be a

useful component of early interventions designed to prevent the onset of internalizing psychopathology following adversity.

The present study was designed to test the proposed framework that the neurobiological underpinnings of emotion regulatory capacity may serve as a buffer against the negative mental health outcomes typically associated with adversity. However, this study should be considered in light of its limitations and the unresolved questions that remain. We focus specifically on outcomes related to early life experiences of threat or maltreatment. Given the highly overlapping experiences of maltreatment and neglect, it is important to make note that children who experience other forms of adversity, such as deprivation or neglect, could also benefit from the protective effects of successful recruitment of regulatory circuitry. Similarly, we have focused on a specific form of cognitive control of emotion—cognitive reappraisal—it is possible that other forms of cognitive control of emotion, such as acceptance of emotional experiences, may modulate negative emotional experiences and the associated amygdala response in a similar way. Future work should investigate the boundary conditions of this model of resilience and determine whether other strategies of emotional regulation may also function as a protective factor buffering children exposed to various forms adversity from development of psychopathology. In addition, the aim of the current study was to determine resilience factors. However, the finding that emotion regulation does not benefit those without a history of maltreatment is interesting and may reflect differential etiological mechanisms of depression in those with and without a history of maltreatment, as proposed by Teicher and colleagues (121). Finally, these analyses identified markers of resilience that were specific to symptoms of depression and not anxiety, which may reflect factors specific to depression (e.g., rumination over past events) that are more readily reframed using reappraisal, as opposed to anxious worries

that have yet to occur. Nonetheless, this divergence provides an important target for future research.

Conclusion

Exposure to child adversity is a potent risk factor for depression and anxiety. Here, we argue that the ability to recruit frontoparietal control networks to modulate amygdala reactivity to negative cues may be a protective factor that buffers children from developing internalizing problems following exposure to adversity. Our findings are consistent with this possibility, demonstrating that children who are more able to modulate amygdala reactivity and recruit prefrontal regions of the frontoparietal network during cognitive reappraisal are less likely to exhibit symptoms of depression following exposure to maltreatment—pointing to a potential neurobiological mechanism of resilience. Greater efforts to identify resilience factors at the neural and behavioral levels can provide mechanistic translational targets for interventions aimed at preventing or treating psychopathology among children who have experienced adversity.

ACCEPTED MANUSCRIPT

Acknowledgements

This work was supported by R01-MH103291 and R01-MH106482 to K.A.M.; NIMH Intramural Research Program ZIAMH002782 to D.S.P.; and NIMH K23MH112872 and Brain and Behavior Research Foundation NARSAD Young Investigator Grant to J.L.J. We are grateful to Debbie Bitran, Andrea Duys, and Azure Reid-Russell for help with participant recruitment and testing, Kelly Sambrook for imaging technical support, and the Stress and Development Lab at Harvard University for helpful discussion.

Disclosures

All authors reported no financial interests or potential conflicts of interest.

Author Contributions

K.A.M., D.S.P., and A.M.R. designed the research; A.M.R. analyzed the data and drafted the manuscript; K.A.M., D.S.P., J.L.J., and D.G.W. provided critical comments and revisions. All authors approved the final version of the manuscript for submission.

ACCEPTED MANUSCRIPT

References

1. Green JG, McLaughlin KA, Berglund PA, Gruber MJ, Sampson NA, Zaslavsky AM, Kessler RC (2010): Childhood Adversities and Adult Psychiatric Disorders in the National Comorbidity Survey Replication I: Associations With First Onset of DSM-IV Disorders. *Arch Gen Psychiatry*. 67: 113–123.

2. Kessler RC, McLaughlin KA, Green JG, Gruber MJ, Sampson NA, Zaslavsky AM, *et al.* (2010): Childhood adversities and adult psychopathology in the WHO World Mental Health Surveys. *British Journal of Psychiatry*. 197: 378–385.

3. McLaughlin KA, Green JG, Gruber MJ, Sampson NA, Zaslavsky AM, Kessler RC (2012): Childhood Adversities and First Onset of Psychiatric Disorders in a National Sample of US Adolescents. *Arch Gen Psychiatry*. 69: 1151–1160.

4. McLaughlin KA (2016): Future Directions in Childhood Adversity and Youth Psychopathology. *Journal of Clinical Child & Adolescent Psychology*. 45: 361–382.

5. Sheridan MA, McLaughlin KA (2014): Dimensions of early experience and neural development: deprivation and threat. *Trends in Cognitive Sciences*. 18: 580–585.

6. McLaughlin KA, Sheridan MA (2016): Beyond Cumulative Risk: A Dimensional Approach to Childhood Adversity. *Curr Dir Psychol Sci*. 25: 239–245.

7. McLaughlin KA, Sheridan MA, Lambert HK (2014): Childhood adversity and neural development: Deprivation and threat as distinct dimensions of early experience. *Neuroscience & Biobehavioral Reviews*. 47: 578–591.

8. Cicchetti D, Toth SL (1995): A Developmental Psychopathology Perspective on Child Abuse and Neglect. *Journal of the American Academy of Child & Adolescent Psychiatry*. 34: 541–565.

9. Cohen P, Brown J, Smailes E (2001): Child abuse and neglect and the development of mental disorders in the general population. *Development and Psychopathology*. 13: 981–999.

10. Lansford JE, Dodge KA, Pettit GS, Bates JE, Crozier J, Kaplow J (2002): A 12-Year Prospective Study of the Long-term Effects of Early Child Physical Maltreatment on Psychological, Behavioral, and Academic Problems in Adolescence. *Arch Pediatr Adolesc Med*. 156: 824–830.

11. Barnes AJ, Lafavor TL, Cutuli JJ, Zhang L, Oberg CN, Masten AS (2017): Health and Self-Regulation among School-Age Children Experiencing Family Homelessness. *Children*. 4: 70.

12. Beardslee WR, Podorefsky D (1988): Resilient adolescents whose parents have serious affective and other psychiatric disorders: importance of self-understanding and relationships. *AJP*. 145: 63–69.

13. Cicchetti D, Rogosch FA, Lynch M, Holt KD (1993): Resilience in maltreated children: Processes leading to adaptive outcome. *Development and Psychopathology*. 5: 629–647.

14. Garmezy N (1993): Children in Poverty: Resilience Despite Risk. *Psychiatry*. 56: 127–136.

15. Masten AS, Best KM, Garmezy N (1990): Resilience and development: Contributions from the study of children who overcome adversity. *Development and Psychopathology*. 2: 425–444.

16. Luthar SS, Cicchetti D, Becker B (2000): The Construct of Resilience: A Critical Evaluation and Guidelines for Future Work. *Child Development*. 71: 543–562.

17. Connor KM, Davidson JRT (2003): Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC). *Depression and Anxiety*. 18: 76–82.

18. Wagnild G (2009): A Review of the Resilience Scale. *Journal of Nursing Measurement*. 17: 105–113.

19. Wagnild G, Young HM (1990): Resilience Among Older Women. *Image: the Journal of Nursing Scholarship*. 22: 252–255.

20. Southwick SM, Charney DS (2012): The Science of Resilience: Implications for the Prevention and Treatment of Depression. *Science*. 338: 79–82.

21. Rutter M (1985): Resilience in the Face of Adversity: Protective Factors and Resistance to Psychiatric Disorder. *The British Journal of Psychiatry*. 147: 598–611.

22. Bronfenbrenner U, Ceci SJ (1994): Nature-nuture reconceptualized in developmental perspective: A bioecological model. *Psychological Review*. 101: 568–586.

23. Ungar M (2011): The Social Ecology of Resilience: Addressing Contextual and Cultural Ambiguity of a Nascent Construct. *American Journal of Orthopsychiatry*. 81: 1–17.

24. Ungar M, Ghazinour M, Richter J (2013): Annual Research Review: What is resilience within the social ecology of human development? *Journal of Child Psychology and Psychiatry*. 54: 348–366.

25. Ungar M (2008): Resilience across Cultures. *Br J Soc Work*. 38: 218–235.

26. Odgers CL, Moffitt TE, Tach LM, Sampson RJ, Taylor A, Matthews CL, Caspi A (2009): The protective effects of neighborhood collective efficacy on British children growing up in deprivation: A developmental analysis. *Developmental Psychology*. 45: 942–957.

27. Benzies K, Mychasiuk R (2009): Fostering family resiliency: a review of the key protective factors. *Child & Family Social Work*. 14: 103–114.

28. Wachs TD (2006): Contributions of Temperament to Buffering and Sensitization Processes in Children's Development. *Annals of the New York Academy of Sciences*. 1094: 28–39.

29. Kim-Cohen J, Gold AL (2009): Measured Gene–Environment Interactions and Mechanisms Promoting Resilient Development. *Curr Dir Psychol Sci*. 18: 138–142.

30. Shechner T, Britton JC, Pérez-Edgar K, Bar-Haim Y, Ernst M, Fox NA, *et al.* (2012): Attention biases, anxiety, and development: toward or away from threats or rewards?: Special Article: Attention Biases, Anxiety, and Development. *Depression and Anxiety*. 29: 282–294.

31. Goldin PR, McRae K, Ramel W, Gross JJ (2008): The Neural Bases of Emotion Regulation: Reappraisal and Suppression of Negative Emotion. *Biological Psychiatry*, Stress, Anxiety, and Post-Traumatic Stress Disorder. 63: 577–586.

32. Gross JJ (1998): The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology*. 2: 271–299.

33. Heller AS, Johnstone T, Shackman AJ, Light SN, Peterson MJ, Kolden GG, *et al.* (2009): Reduced capacity to sustain positive emotion in major depression reflects diminished maintenance of fronto-striatal brain activation. *PNAS*. 106: 22445–22450.

34. Ochsner KN, Bunge SA, Gross JJ, Gabrieli JDE (2002): Rethinking Feelings: An fMRI Study of the Cognitive Regulation of Emotion. *Journal of Cognitive Neuroscience*. 14: 1215–1229.

35. Ochsner KN, Ray RD, Cooper JC, Robertson ER, Chopra S, Gabrieli JDE, Gross JJ (2004): For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage*. 23: 483–499.

36. Silvers JA, Shu J, Hubbard AD, Weber J, Ochsner KN (2015): Concurrent and lasting effects of emotion regulation on amygdala response in adolescence and young adulthood. *Developmental Science*. 18: 771–784.

37. Gross JJ (1998): Antecedent- and Response-Focused Emotion Regulation: Divergent Consequences for Experience, Expression, and Physiology. 14.

38. Gross JJ, John OP (2003): Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*. 85: 348–362.

39. McRae K, Hughes B, Chopra S, Gabrieli JDE, Gross JJ, Ochsner KN (2009): The Neural Bases of Distraction and Reappraisal. *Journal of Cognitive Neuroscience*. 22: 248–262.

40. Webb TL, Miles E, Sheeran P (2012): Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin*. 138: 775–808.

41. John OP, Gross JJ (2004): Healthy and Unhealthy Emotion Regulation: Personality Processes, Individual Differences, and Life Span Development. *Journal of Personality*. 72: 1301–1334.

42. Beck AT (1979): *Cognitive Therapy of Depression*. Guilford Press.

43. Chorpita BF, Daleiden EL (2009): Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology*. 77: 566–579.

44. Compton SN, March JS, Brent D, Albano AM, Weersing VR, Curry J (2004): Cognitive-Behavioral Psychotherapy for Anxiety and Depressive Disorders in Children and Adolescents: An Evidence-Based Medicine Review. *Journal of the American Academy of Child & Adolescent Psychiatry*. 43: 930–959.

45. Ochsner KN, Gross JJ (2005): The cognitive control of emotion. *Trends in Cognitive Sciences*. 9: 242–249.

46. Davis JI, Gross JJ, Ochsner KN (2011): Psychological distance and emotional experience: What you see is what you get. *Emotion*. 11: 438–444.

47. Denny BT, Ochsner KN (2014): Behavioral effects of longitudinal training in cognitive reappraisal. *Emotion*. 14: 425–433.

48. Erk S, Mikschl A, Stier S, Ciaramidaro A, Gapp V, Weber B, Walter H (2010): Acute and Sustained Effects of Cognitive Emotion Regulation in Major Depression. *J Neurosci*. 30: 15726–15734.

49. McLaughlin KA, Peverill M, Gold AL, Alves S, Sheridan MA (2015): Child Maltreatment and Neural Systems Underlying Emotion Regulation. *J Am Acad Child Adolesc Psychiatry*. 54: 753–762.

50. Buhle JT, Silvers JA, Wager TD, Lopez R, Onyemekwu C, Kober H, *et al.* (2014): Cognitive Reappraisal of Emotion: A Meta-Analysis of Human Neuroimaging Studies. *Cerebral Cortex*. 24: 2981–2990.

51. Miller EK, Cohen JD (2001): An Integrative Theory of Prefrontal Cortex Function. *Annu Rev Neurosci*. 24: 167–202.

52. Niendam TA, Laird AR, Ray KL, Dean YM, Glahn DC, Carter CS (2012): Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn Affect Behav Neurosci*. 12: 241–268.

53. Banks SJ, Eddy KT, Angstadt M, Nathan PJ, Phan KL (2007): Amygdala–frontal connectivity during emotion regulation. *Soc Cogn Affect Neurosci*. 2: 303–312.

54. Eippert F, Veit R, Weiskopf N, Erb M, Birbaumer N, Anders S (2007): Regulation of emotional responses elicited by threat-related stimuli. *Human Brain Mapping*. 28: 409–423.

55. Lee H, Heller AS, van Reekum CM, Nelson B, Davidson RJ (2012): Amygdala–prefrontal coupling underlies individual differences in emotion regulation. *NeuroImage*. 62: 1575–1581.

56. Ochsner KN, Silvers JA, Buhle JT (2012): Functional imaging studies of emotion regulation: a synthetic review and evolving model of the cognitive control of emotion. *Annals of the New York Academy of Sciences*. 1251: E1–E24.

57. Silvers JA, Insel C, Powers A, Franz P, Helion C, Martin RE, *et al.* (2017): vlPFC–vmPFC–Amygdala Interactions Underlie Age-Related Differences in Cognitive Regulation of Emotion. *Cereb Cortex*. 27: 3502–3514.

58. McRae K, Gross JJ, Weber J, Robertson ER, Sokol-Hessner P, Ray RD, *et al.* (2012): The development of emotion regulation: an fMRI study of cognitive reappraisal in children, adolescents and young adults. *Social cognitive and affective neuroscience*. 7: 11–22.

59. Silvers JA, McRae K, Gabrieli JDE, Gross JJ, Remy KA, Ochsner KN (2012): Age-related differences in emotional reactivity, regulation, and rejection sensitivity in adolescence. *Emotion*. 12: 1235–1247.

60. Ahmed SP, Somerville LH, Sebastian CL (2018): Using temporal distancing to regulate emotion in adolescence: modulation by reactive aggression. *Cognition and Emotion*. 32: 812–826.

61. Van Cauwenberge V, Van Leeuwen K, Hoppenbrouwers K, Wiersema JR (2017): Developmental changes in neural correlates of cognitive reappraisal: An ERP study using the late positive potential. *Neuropsychologia*. 95: 94–100.

62. Levesque J, Joanette Y, Mensour B, Beaudoin G, Leroux J-M, Bourgouin P, Beauregard M (2004): Neural basis of emotional self-regulation in childhood. *Neuroscience*. 129: 361–369.

63. McRae K, Misra S, Prasad AK, Pereira SC, Gross JJ (2012): Bottom-up and top-down emotion generation: implications for emotion regulation. *Soc Cogn Affect Neurosci*. 7: 253–262.

64. Pitskel NB, Bolling DZ, Kaiser MD, Crowley MJ, Pelphrey KA (2011): How grossed out are you? The neural bases of emotion regulation from childhood to adolescence. *Developmental Cognitive Neuroscience*. 1: 324–337.

65. Anderson AK, Phelps EA (2001): Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*. 411: 305–309.

66. Cunningham WA, Brosch T (2012): Motivational Salience: Amygdala Tuning From Traits, Needs, Values, and Goals. *Current Directions in Psychological Science*. 21: 54–59.

67. Vuilleumier P, Pourtois G (2007): Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging. *Neuropsychologia*, The Perception of Emotion and Social Cues in Faces. 45: 174–194.

68. Hartley CA, Phelps EA (2010): Changing Fear: The Neurocircuitry of Emotion Regulation. *Neuropsychopharmacology*. 35: 136–146.

69. Johnstone T, Reekum CM van, Urry HL, Kalin NH, Davidson RJ (2007): Failure to Regulate: Counterproductive Recruitment of Top-Down Prefrontal-Subcortical Circuitry in Major Depression. *J Neurosci*. 27: 8877–8884.

70. Phelps EA, Delgado MR, Nearing KI, LeDoux JE (2004): Extinction Learning in Humans: Role of the Amygdala and vmPFC. *Neuron*. 43: 897–905.

71. Quirk GJ, Likhtik E, Pelletier JG, Paré D (2003): Stimulation of Medial Prefrontal Cortex Decreases the Responsiveness of Central Amygdala Output Neurons. *J Neurosci*. 23: 8800–8807.

72. Urry HL, Reekum CM van, Johnstone T, Kalin NH, Thurow ME, Schaefer HS, *et al.* (2006): Amygdala and Ventromedial Prefrontal Cortex Are Inversely Coupled during Regulation of Negative Affect and Predict the Diurnal Pattern of Cortisol Secretion among Older Adults. *J Neurosci*. 26: 4415–4425.

73. Beauregard M, Paquette V, Levesque J (2006): Dysfunction in the neural circuitry of emotional self-regulation in major depressive disorder. *NeuroReport.* 17: 843.

74. Carthy T, Horesh N, Apter A, Edge MD, Gross JJ (2010): Emotional reactivity and cognitive regulation in anxious children. *Behaviour Research and Therapy*. 48: 384–393.

75. Goldin PR, Manber T, Hakimi S, Canli T, Gross JJ (2009): Neural Bases of Social Anxiety Disorder: Emotional Reactivity and Cognitive Regulation During Social and Physical Threat. *Arch Gen Psychiatry*. 66: 170–180.

76. Miller AB, McLaughlin KA, Busso DS, Brueck S, Peverill M, Sheridan MA (2018): Neural Correlates of Emotion Regulation and Adolescent Suicidal Ideation. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 3: 125–132.

77. Perlman G, Simmons AN, Wu J, Hahn KS, Tapert SF, Max JE, *et al.* (2012): Amygdala response and functional connectivity during emotion regulation: A study of 14 depressed adolescents. *Journal of Affective Disorders*. 139: 75–84.

78. Pico-Perez M, Radua J, Steward T, Menchón JM, Soriano-Mas C (2017): Emotion regulation in mood and anxiety disorders: A meta-analysis of fMRI cognitive reappraisal studies. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. 79: 96–104.

79. Glaser JP, van Os J, Portegijs PJM, Myin-Germeys I (2006): Childhood trauma and emotional reactivity to daily life stress in adult frequent attenders of general practitioners. *Journal of Psychosomatic Research*. 61: 229–236.

80. Heleniak C, Jenness JL, Vander Stoep A, McCauley E, McLaughlin KA (2016): Childhood Maltreatment Exposure and Disruptions in Emotion Regulation: A Transdiagnostic Pathway to Adolescent Internalizing and Externalizing Psychopathology. *Cogn Ther Res*. 40: 394–415.

81. Weissman DG, Bitran D, Miller AB, Schaefer JD, Sheridan MA, McLaughlin KA (in press): Disruptions in emotion regulation as a transdiagnostic mechanism linking child maltreatment with the emergence of psychopathology. *Development and Psychopathology*. .

82. Wichers M, Schrijvers D, Geschwind N, Jacobs N, Myin-Germeys I, Thiery E, *et al.* (2009): Mechanisms of gene–environment interactions in depression: evidence that genes potentiate multiple sources of adversity. *Psychological Medicine*. 39: 1077–1086.

83. Heleniak C, McLaughlin KA, Ormel J, Riese H (2016): Cardiovascular reactivity as a mechanism linking child trauma to adolescent psychopathology. *Biological Psychology*. 120: 108–119.

84. McLaughlin KA, Sheridan MA, Alves S, Mendes WB (2014): Child Maltreatment and Autonomic Nervous System Reactivity: Identifying Dysregulated Stress Reactivity Patterns by Using the Biopsychosocial Model of Challenge and Threat. *Psychosomatic Medicine*. 76: 538–546.

85. Herringa RJ, Birn RM, Ruttle PL, Burghy CA, Stodola DE, Davidson RJ, Essex MJ (2013): Childhood maltreatment is associated with altered fear circuitry and increased internalizing symptoms by late adolescence. *PNAS*. 110: 19119–19124.

86. Marusak HA, Martin KR, Etkin A, Thomason ME (2015): Childhood Trauma Exposure Disrupts the Automatic Regulation of Emotional Processing. *Neuropsychopharmacology*. 40: 1250–1258.

87. McCrory EJ, Brito SAD, Kelly PA, Bird G, Sebastian CL, Mechelli A, *et al.* (2013): Amygdala activation in maltreated children during pre-attentive emotional processing. *The British Journal of Psychiatry*. 202: 269–276.

88. McCrory EJ, De Brito SA, Viding E (2011): The Impact of Childhood Maltreatment: A Review of Neurobiological and Genetic Factors. *Front Psychiatry*. 2. doi: 10.3389/fpsyt.2011.00048.

89. Tottenham N, Hare TA, Millner A, Gilhooly T, Zevin J, Casey BJ (2011): Elevated Amygdala Response to Faces Following Early Deprivation. *Dev Sci*. 14: 190–204.

90. Hein TC, Monk CS (2017): Research Review: Neural response to threat in children, adolescents, and adults after child maltreatment – a quantitative meta-analysis. *Journal of Child Psychology and Psychiatry*. 58: 222–230.

91. McLaughlin KA, Lambert HK (2017): Child trauma exposure and psychopathology: mechanisms of risk and resilience. *Current Opinion in Psychology*, Traumatic stress. 14: 29–34.

92. Anthony JL, Lonigan CJ, Hooe ES, Phillips BM (2002): An Affect-Based, Hierarchical Model of Temperament and Its Relations With Internalizing Symptomatology. *Journal of Clinical Child & Adolescent Psychology*. 31: 480–490.

93. Heleniak C, King KM, Monahan KC, McLaughlin KA (2018): Disruptions in Emotion Regulation as a Mechanism Linking Community Violence Exposure to Adolescent Internalizing Problems. *Journal of Research on Adolescence*. 28: 229–244.

94. McLaughlin KA, Hatzenbuehler ML, Hilt LM (2009): Emotion dysregulation as a mechanism linking peer victimization to internalizing symptoms in adolescents. *Journal of Consulting and Clinical Psychology*. 77: 894–904.

95. Silk JS, Steinberg L, Morris AS (2003): Adolescents' Emotion Regulation in Daily Life: Links to Depressive Symptoms and Problem Behavior. *Child Development*. 74: 1869–1880.

96. Kim J, Cicchetti D (2010): Longitudinal pathways linking child maltreatment, emotion regulation, peer relations, and psychopathology. *J Child Psychol Psychiatry*. 51: 706–716.

97. Gee DG, Gabard-Durnam LJ, Flannery J, Goff B, Humphreys KL, Telzer EH, *et al.* (2013): Early developmental emergence of human amygdala–prefrontal connectivity after maternal deprivation. *PNAS*. 201307893.

98. Hanson JL, Chung MK, Avants BB, Shirtcliff EA, Gee JC, Davidson RJ, Pollak SD (2010): Early Stress Is Associated with Alterations in the Orbitofrontal Cortex: A Tensor-Based Morphometry Investigation of Brain Structure and Behavioral Risk. *J Neurosci*. 30: 7466–7472.

99. Herringa RJ, Burghy CA, Stodola DE, Fox ME, Davidson RJ, Essex MJ (2016): Enhanced prefrontal-amygdala connectivity following childhood adversity as a protective mechanism against internalizing in adolescence. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 1: 326–334.

100. Miller GE, Chen E, Armstrong CC, Carroll AL, Ozturk S, Rydland KJ, *et al.* (2018): Functional connectivity in central executive network protects youth against cardiometabolic risks linked with neighborhood violence. *Proceedings of the National Academy of Sciences*. 6.

101. McLaughlin KA, Conron KJ, Koenen KC, Gilman SE (2010): Childhood adversity, adult stressful life events, and risk of past-year psychiatric disorder: a test of the stress sensitization hypothesis in a population-based sample of adults. *Psychological Medicine*. 40: 1647–1658.

102. Hammen C (1991): Generation of stress in the course of unipolar depression. *Journal of Abnormal Psychology*. 100: 555–561.

103. Hammen C (2006): Stress generation in depression: Reflections on origins, research, and future directions. *Journal of Clinical Psychology*. 62: 1065–1082.

104. Kendler KS, Karkowski LM, Prescott CA (1999): Causal Relationship Between Stressful Life Events and the Onset of Major Depression. *AJP*. 156: 837–841.

105. Mazure CM (1998): Life Stressors as Risk Factors in Depression. *Clinical Psychology: Science and Practice*. 5: 291–313.

106. Bandoli G, Campbell-Sills L, Kessler RC, Heeringa SG, Nock MK, Rosellini AJ, *et al.* (2017): Childhood adversity, adult stress, and the risk of major depression or generalized anxiety disorder in US soldiers: a test of the stress sensitization hypothesis. *Psychological Medicine*. 47: 2379–2392.

107. Espejo EP, Hammen CL, Connolly NP, Brennan PA, Najman JM, Bor W (2007): Stress Sensitization and Adolescent Depressive Severity as a Function of Childhood Adversity: A Link to Anxiety Disorders. *J Abnorm Child Psychol*. 35: 287–299.

108. Hammen C, Henry R, Daley SE (2000): Depression and sensitization to stressors among young women as a function of childhood adversity. *Journal of Consulting and Clinical Psychology*. 68: 782–787.

109. Kendler KS, Kuhn JW, Prescott CA (2004): Childhood sexual abuse, stressful life events and risk for major depression in women. *Psychological Medicine*. 34: 1475–1482.

110. Troy AS, Wilhelm FH, Shallcross AJ, Mauss IB (2010): Seeing the Silver Lining: Cognitive Reappraisal Ability Moderates the Relationship Between Stress and Depressive Symptoms. *Emotion*. 10: 783–795.

111. Lyons DM, Parker KJ, Schatzberg AF (2010): Animal models of early life stress: Implications for understanding resilience. *Developmental Psychobiology*. 52: 616–624.

112. Badura-Brack AS, Naim R, Ryan TJ, Levy O, Abend R, Khanna MM, *et al.* (2015): Effect of Attention Training on Attention Bias Variability and PTSD Symptoms: Randomized Controlled Trials in Israeli and U.S. Combat Veterans. *American Journal of Psychiatry*. 172: 1233–1241.

113. Lazarov A, Marom S, Yahalom N, Pine DS, Hermesh H, Bar-Haim Y (2018): Attention bias modification augments cognitive–behavioral group therapy for social anxiety disorder: a randomized controlled trial. *Psychological Medicine*. 48: 2177–2185.

114. Wald I, Fruchter E, Ginat K, Stolin E, Dagan D, Bliese PD, *et al.* (2016): Selective prevention of combat-related post-traumatic stress disorder using attention bias modification training: a randomized controlled trial. *Psychological Medicine*. 46: 2627–2636.

115. White LK, Sequeira S, Britton JC, Brotman MA, Gold AL, Berman E, *et al.* (2017): Complementary Features of Attention Bias Modification Therapy and Cognitive-

Behavioral Therapy in Pediatric Anxiety Disorders. *American Journal of Psychiatry*. 174: 775–784.

116. MacLeod C, Rutherford E, Campbell L, Ebsworthy G, Holker L (2002): Selective attention and emotional vulnerability: Assessing the causal basis of their association through the experimental manipulation of attentional bias. *Journal of Abnormal Psychology*. 111: 107–123.

117. Bar□Haim Y (2010): Research Review: attention bias modification (ABM): a novel treatment for anxiety disorders. *Journal of Child Psychology and Psychiatry*. 51: 859–870.

118. Briggs□Gowan MJ, Grasso D, Bar□Haim Y, Voss J, McCarthy KJ, Pine DS, Wakschlag LS (2016): Attention bias in the developmental unfolding of post-traumatic stress symptoms in young children at risk. *Journal of Child Psychology and Psychiatry*. 57: 1083–1091.

119. Clark DA, Beck AT (2010): Cognitive theory and therapy of anxiety and depression: Convergence with neurobiological findings. *Trends in Cognitive Sciences*. 14: 418–424.

120. Frewen PA, Dozois DJA, Lanius RA (2008): Neuroimaging studies of psychological interventions for mood and anxiety disorders: Empirical and methodological review. *Clinical Psychology Review*. 28: 228–246.

121. Teicher MH, Samson JA (2013): Childhood Maltreatment and Psychopathology: A Case for Ecophenotypic Variants as Clinically and Neurobiologically Distinct Subtypes. *AJP*. 170: 1114–1133.

Figure and Table Legends

Table 1. Summary of Participant Demographic and Clinical Characteristics

Figure 1. Schematic representation of experimental task. While undergoing fMRI, participants were instructed to either look at an emotional image (neutral or negative) or attempt to decrease their emotional response using cognitive reappraisal (negative images only). Following the presentation of the image, participants reported the strength of emotion they experienced while viewing the image. Between trials, participants were instructed to relax.

Figure 2. Amygdala activity during cognitive reappraisal moderates the association between child maltreatment and depression symptoms over time. A. Red mask overlay shows left amygdala ROI. B. Greater reduction in amygdala activation during cognitive reappraisal relative to passive viewing of negative emotional stimuli (decrease negative > look negative) is associated with lower depression symptoms over time among children who were maltreated (blue) but is unrelated to the development of depression symptoms among those without a history of maltreatment (black). Shaded region indicates 95% CI.

Figure 3. Association of child maltreatment with depression is moderated by prefrontal recruitment during cognitive reappraisal. A. Red mask overlay shows right superior frontal gyrus (SFG) ROI. B. Greater recruitment of the right SFG during cognitive reappraisal relative to passive viewing of negative emotional stimuli (decrease negative > look negative) is associated with reduced symptoms of depression among children exposed to maltreatment (blue) but is unrelated to depression symptoms among those without a history of maltreatment (black). C. Red mask overlay shows right dorsal anterior cingulate cortex (dACC) ROI. D. Greater recruitment of the right dACC during cognitive reappraisal relative to passive viewing of negative emotional stimuli (decrease negative > look negative) is associated with reduced symptoms of depression among children exposed to maltreatment (blue) but is unrelated to depression symptoms among those without a history of maltreatment (black). Shaded region indicates 95% CI.

Figure 4. Association between reported tendency to engage in cognitive reappraisal in daily life and depression is moderated by severity of abuse history. Greater reported use of cognitive reappraisal strategies in response to stressful life events is associated with lower symptoms of depression, particularly among children with more severe maltreatment history (blue) compared to those with a less severe or no history of maltreatment (black). Reappraisal use measured using the Emotion Regulation Questionnaire (ERQ) (38). Shaded region indicates 95% CI.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Participant Characteristics	**Non-maltreated (N=72)**		**Maltreated (N=79)**			
	Mean	*SD*	*Mean*	*SD*	*t-value*	*p-value*
Age	12.57	2.55	12.94	2.69	-0.86	0.394
Pubertal Stage	2.88	0.74	3.13	0.86	-1.8791	0.062
CDI at Baseline	5.54	4.27	12.16	8.72	-6.01	<.001
CDI at Follow Up	6.14	5.33	10.63	8.41	-3.37	<.001
SCARED at Baseline	26.51	15.70	15.26	10.52	-5.17	<.001
SCARED at Follow Up	21.58	15.12	16.72	9.43	-2.03	0.045
ERQ - Reappraisal	27.46	6.34	27.32	8.15	0.12	0.906
	N	*%*	*N*	*%*	X^2	*p-value*
Sex: Female	33	45.83%	43	54.43%	0.80	0.372
Attrition	15	20.83%	24	30.38%	1.33	0.249
Race: Non-White	22	30.56%	61	77.22%	31.27	<.001
Adversity Exposure						
Physical Abuse	0	0%	57	72%	80.41	<.001
Emotional Abuse	3	4%	44	56%	44.28	<.001
Sexual Abuse	0	0%	36	46%	40.61	<.001
Neglect	0	0%	19	24%	17.68	<.001

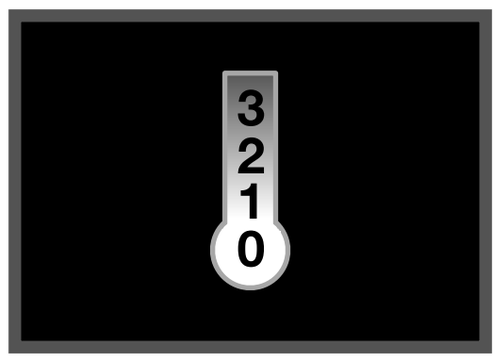
Instructional Cue
2 secs



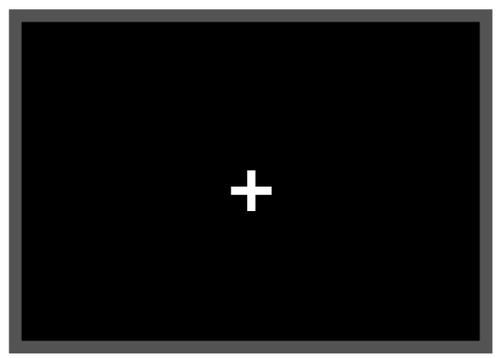
Passive Viewing or Regulation
6-10 secs



Affect Rating
2 secs



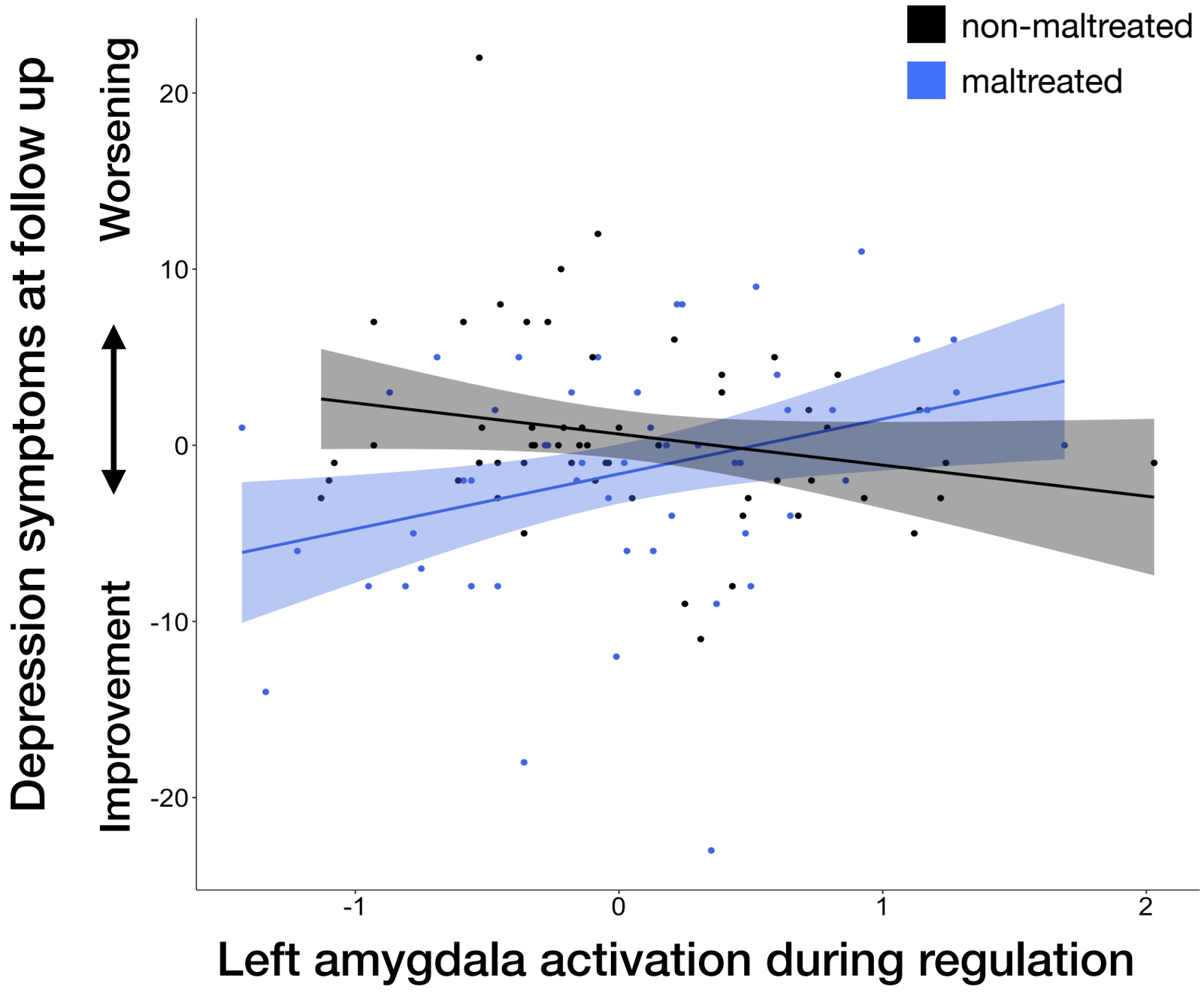
Relax
2-6 secs



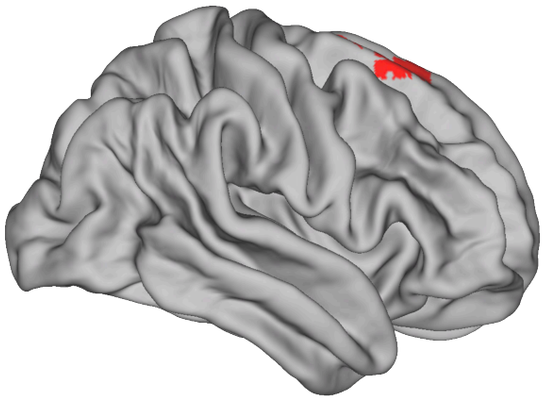
A.



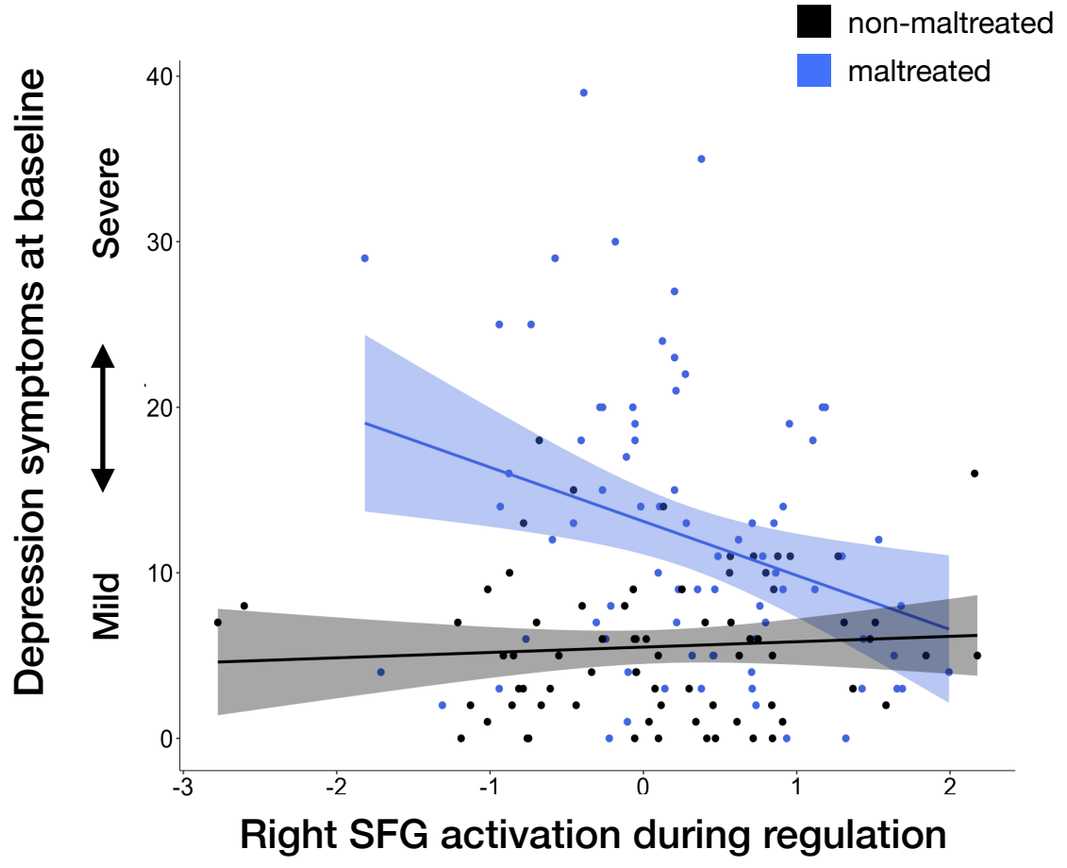
B.


non-maltreated
maltreated
Depression symptoms at follow up
Worsening
Improvement
20
10
0
-10
-20
-1
0
1
2
Left amygdala activation during regulation

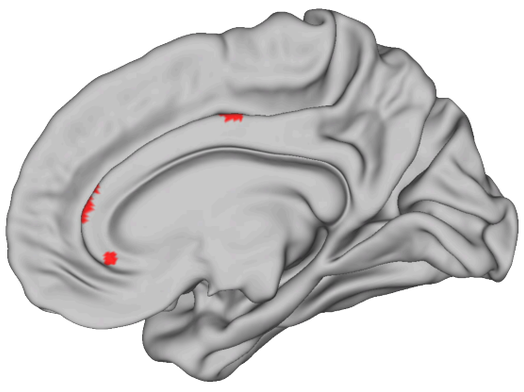
A.



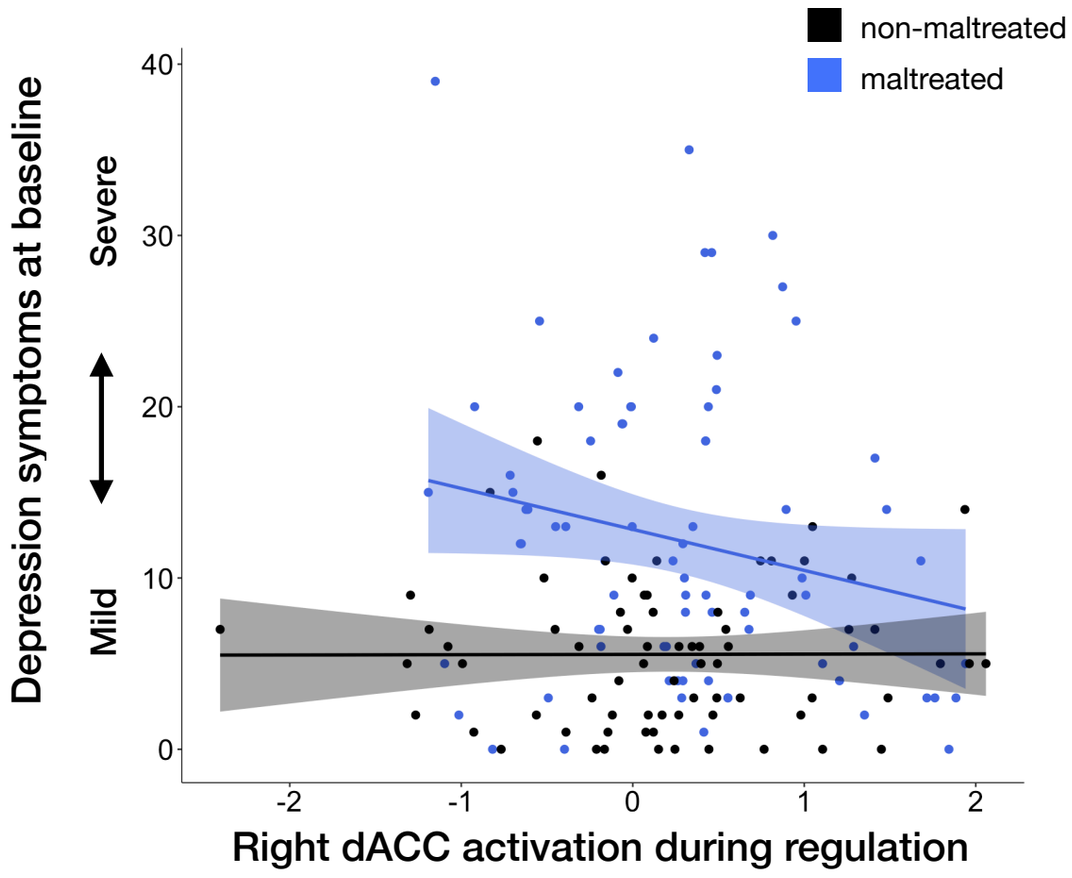
B.

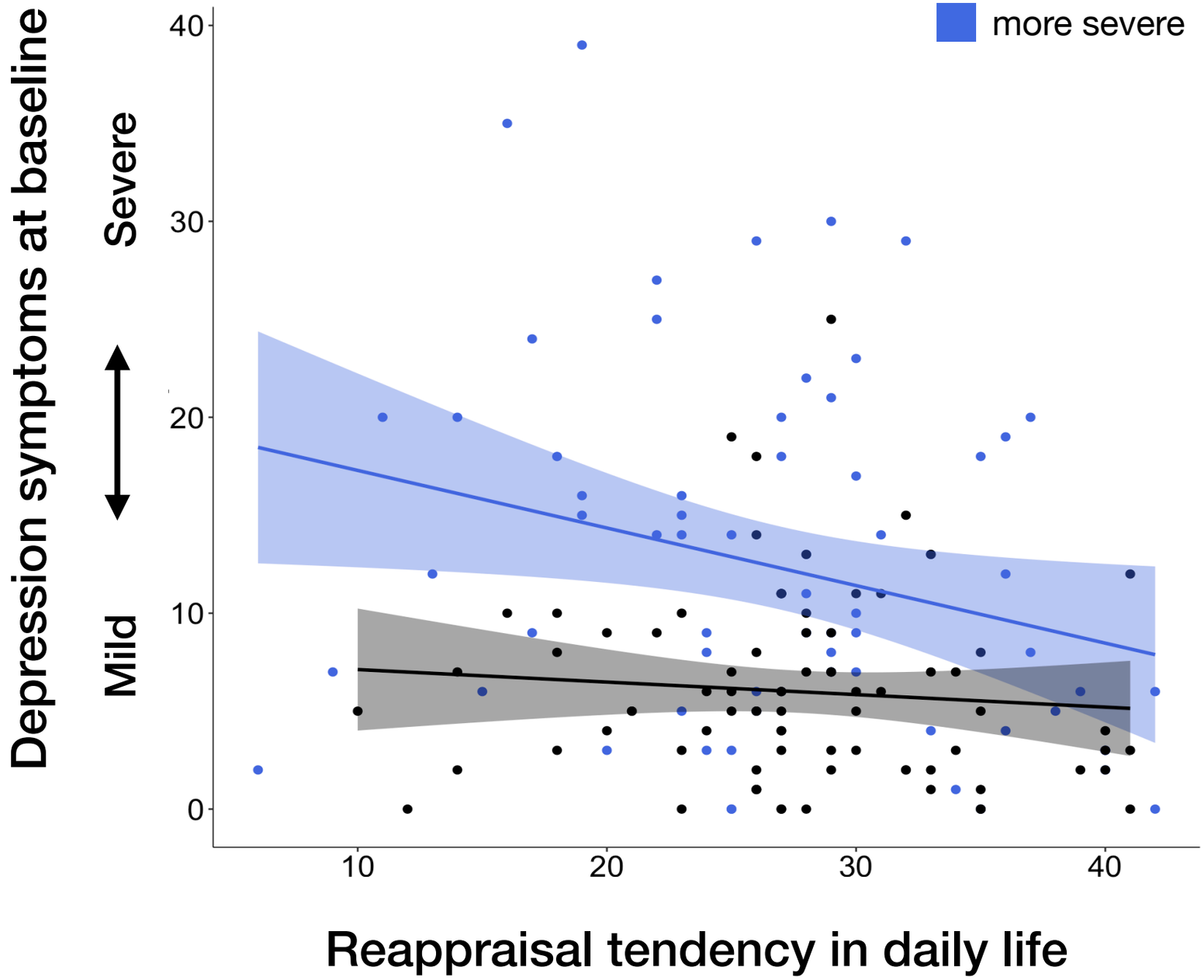

non-maltreated
maltreated
Depression symptoms at baseline
Severe
Mild
40
30
20
10
0
-3
-2
-1
0
1
2
Right SFG activation during regulation

C.



D.


non-maltreated
maltreated
Depression symptoms at baseline
Severe
Mild
40
30
20
10
0
-2
-1
0
1
2
Right dACC activation during regulation


Abuse severity:
less severe
more severe
Depression symptoms at baseline
Severe
Mild
40
30
20
10
0
10
20
30
40
Reappraisal tendency in daily life

ACCEPTED MANUSCRIPT

Neurobiological Markers of Resilience to Depression Following Childhood Maltreatment: The Role of Neural Circuits Supporting the Cognitive Control of Emotion

Supplemental Information

Supplemental Methods

Participants

160 participants aged 8-17 (M=12.76, SD=2.62) drawn from a larger parent study were recruited from schools, after-school and prevention programs, adoption programs, food banks, shelters, parenting programs, medical clinics, and the general community in Seattle, WA between January 2015 and June 2017. Participants were recruited based on inclusion criteria for the maltreated group, including exposure to physical or sexual abuse or direct witnessing of domestic violence (i.e., violence directed towards a caregiver), while participants in the control group were required to have an absence of exposure to maltreatment or other forms of interpersonal violence and were matched to children in the maltreated group on age, sex, and handedness. Exclusion criteria included IQ < 80, pervasive developmental disorder, active psychotic symptoms, mania, substance abuse, other safety concerns, as well as any standard contraindications for MRI scanning (e.g., braces). Approximately half of the sample was female (50.3%) and 45.0% of the sample identified as White, 23.2% as Black, 10.6% as Hispanic/Latino, 11.3% as Asian or Pacific Islander, and 9.9% as biracial or other racial/ethnic groups.

Of the 160 children who initially completed the neuroimaging assessment, 9 were excluded from analysis due to either: excessive motion during scanning (n=2), lack of behavioral responding in more than 2 runs of the emotion regulation task (n=2), failure to complete the task (n=4), and exposure to significant violence that was not maltreatment-related (n=1). All procedures were

ACCEPTED MANUSCRIPT

approved by the Institutional Review Board at the University of Washington. Written informed consent was obtained from legal guardians while children provided written assent.

Measures

History of Maltreatment. Children and parents were assessed for history of exposure to emotional, sexual, or physical abuse using a multi-informant, multi-method approach to determine participants' exposure to maltreatment. Children were classified as experiencing emotional, physical, or sexual abuse if it was endorsed by the child or parent on the following: Childhood Experiences of Care and Abuse (CECA) interview (1), UCLA PTSD Reaction Index (PTSD-RI) trauma screen (2), Violence Exposure Scale for Children-Revised (VEX-R) (3), Childhood Trauma Questionnaire (CTQ) (4, 5), the Juvenile Victimization Questionnaire (JVQ) (6), or PTSD-RI trauma screen. Domestic violence was assessed by child-report only on the CECA interview and PTSD-RI. Validated thresholds for exposure to physical and sexual abuse (5) were applied to determine abuse exposure based on the CTQ. Children were classified as experiencing physical or sexual abuse if abuse was endorsed by the child (on the CECA interview, PTSD-RI trauma screen, or above the validated CTQ threshold) or parent (on the JVQ or PTSD-RI trauma screen). A total of 79 children (51.7%) experienced maltreatment, of whom 46% experienced sexual abuse, 72% physical abuse, and 56% emotional abuse. While the recruitment focus of this sample was to recruit youth exposed to threat-related adversity in the form of violence, other types of adversity, such as neglect, commonly co-occur with abuse and this polyvictimization was reflected in our sample with 24% experiencing neglect. Inter-rater reliability was good for child and caregiver maltreatment reports (82.0% agreement; kappa=0.62). A maltreatment severity score was created by summing the total number of types of maltreatment and violence exposure experienced by the

ACCEPTED MANUSCRIPT

child, including physical abuse, sexual abuse, emotional abuse, domestic violence, and exposure to violence in the broader community. Each type of maltreatment or violence exposure was coded as present or absent, and these indicators were summed to create a severity score.

Depression Symptoms. Symptoms of depression were measured using the Children's Depression Inventory (CDI) (7). This self-report questionnaire is a 27-item measure that is widely used and validated for use in children aged 8-17. Each question probes a specific symptom that is answered at three levels of severity (i.e., "once in a while," "many times," "all the time"). These items were summed for a total score at baseline and at the follow-up visit approximately two years later (M=20.33 months, SD=6.98 months). Similar to previous studies of reliability and validity (7), the CDI demonstrated good internal consistency in the current sample (Cronbach's alpha=.89).

Anxiety Symptoms. Symptoms of anxiety were measured using the Screen for Child Anxiety Related Emotional Disorders (SCARED) (8), which measures anxiety symptoms across five domains: panic/somatic, generalized anxiety, separation anxiety, social phobia, and school phobia. The SCARED has good psychometric properties (8) and had excellent internal consistency in the current sample (Cronbach's alpha = .94). This measure was collected at baseline and at the follow-up visit.

Emotion Regulation Strategies. Tendency to employ reappraisal in everyday life following stressful events was assessed using the Emotion Regulation Questionnaire (ERQ) (9). The ERQ is a widely used 10-item self-report questionnaire, with subscales for specific emotion regulation strategies (e.g., reappraisal). Answers are given on a seven-point Likert scale, from "strongly disagree" to "strongly agree." The internal consistency was good in the current sample (Cronbach's alpha = .84).

ACCEPTED MANUSCRIPT

Procedure

Emotion Regulation Task. Participants completed an experimental task designed to evoke cognitive reappraisal which has been widely used in adults (10) and adapted for use in pediatric samples (11, 12). Age-appropriate negative and neutral images from a stimulus set developed in the lab of the senior author (KAM) were used (images and normative ratings are publicly available: https://osf.io/43hfq/). Participants were trained on the emotion regulation task at a previous study visit and instructions were reviewed in the scanner. For 'Look' trials participants were trained to attend to the picture and to allow their emotional reaction to unfold naturally. During 'Far' trials (hereafter referred to as Regulation) participants were trained to modulate their emotional response to negative images using cognitive reappraisal. Participants were taught two reappraisal strategies prior to the scan: psychological distancing (e.g. "imagine the situation as far away from you,") and reinterpretation (e.g. "imagine the situation is fake or not real," or "pretend the people in the picture are just actors in a movie"). Following image presentation participants were asked to rate the intensity of their emotional response on a 'feelings thermometer' ranging from 0 to 3. Length of stimulus presentation (6-10 seconds) and inter-trial interval (ITI) were jittered (2-6 seconds). Participants completed 4 runs of the task, each run consisting of 5 Look Neutral, 5 Look Negative, and 5 Regulation (i.e., Far Negative) trials, resulting in 20 trials of each condition across the 4 runs of the task. Negative images in Look and Regulation trials were matched on valence and arousal. Trials were presented in random order within each run.

Image Acquisition and Preprocessing. All images were acquired using a Phillips Achieva 3T scanner using a 32-channel head coil at the University of Washington's Integrated Brain Imaging Center. Anatomical scans (T1-weighted MPRAGE volumes; TR=2530ms, TE=3.5ms, flip angle=7°, FOV=256×256, 176 slices, in-plane voxel size=1mm^3) were acquired for co-registration

ACCEPTED MANUSCRIPT

with functional magnetic resonance imaging (fMRI). Blood oxygenation level dependent (BOLD) signal during functional runs was acquired using a gradient-echo T2*-weighted echo planar imaging (EPI) sequence. 37 3mm thick axial slices were acquired sequentially and parallel to the AC-PC line (TR=2s, TE=25ms, flip angle=79°, Inter-slice gap=.6mm, FOV=224×224×132.6, matrix size=76x74). Prior to each scan, four images were acquired and discarded to allow longitudinal magnetization to reach equilibrium.

Data preprocessing consisted of the following steps: motion correction with FSL MCFLIRT, slice timing correction with FSL slicetimer, despiking with AFNI 3dDespike, and spatial smoothing with FSL SUSAN (using a Gaussian kernel size of 6-mm FWHM) (13–15). Nuisance regressors entered into person-level models consisted of 6 rigid-body motion regressors as well as time-series extracted from white matter and ventricles entered to control for physiological noise (16). Outlier volumes in which framewise displacement exceeded 1 mm, the derivative of variance in BOLD signal across the brain (DVARS) exceeded the upper fence (above 75th percentile + 1.5 x inter-quartile range), or signal intensity was more than 3 SD from the mean were excluded by regressing these volumes out of first-level models. This resulted in a total of 285 frame displacements (M=0.181, SD=0.153, average of 2.22% frames) for the non-maltreatment group and 310 frame displacements (M=0.240, SD=0.219, average of 3.93% frames) for the maltreatment group, which was not significantly different (t(149)=1.901, p=.060). Following estimation of the individual-level models, the resulting contrast images were registered first to a custom template, representing the average anatomy of the sample, and then from the custom template to the standard space of the Montreal Neurological Institute (MNI) template. Anatomical co-registration of the functional data with each participant's T1-weighted image and normalization were performed using Advanced Normalization Tools (ANTs) (17).

ACCEPTED MANUSCRIPT

Statistical Analysis

fMRI. Event-related regressors were created by convolving a boxcar function of phase, duration, and amplitude one with the standard (double-gamma) hemodynamic response function for each phase of the task (i.e., instruction, stimulus, rating). Regressors for stimulus presentation were separated by trial type (Look Neutral, Look Negative, Regulation). A general linear model was constructed for each participant and the two contrasts used in prior studies of this task (19–21) were examined: Look Negative > Neutral (i.e., emotional reactivity), and the Regulation > Look Negative (i.e., emotion regulation). Individual-level estimates of BOLD activity were submitted to group-level random effects models using FSL's FLAME 1 (21) with cluster-level correction of $z > 2.3$, $p < .01$. We examined differences in BOLD response during contrasts of interest in whole-brain analyses to confirm expected effects of the experimental task (see Supplemental Table S1).

Amygdala ROI Analysis. We examined amygdala activation using a region-of-interest (ROI) analysis, separately for left and right amygdala. Functional ROIs were created by intersecting a structural mask of the amygdala from the Harvard-Oxford subcortical atlas (20% threshold) with the map of activation in the entire sample for each contrast. Linear regression models were computed with depression symptoms at baseline and follow up as the outcome measure, amygdala activation during regulation, history of trauma and their interactions as predictors, while controlling for socioeconomic status (i.e., income to needs ratio), non-white status, sex, and age. When examining follow up depression as the outcome of interest, days between baseline and follow up and baseline symptoms of depression were also included as covariates. Analyses were repeated for anxiety symptoms.

ACCEPTED MANUSCRIPT

Exploratory Prefrontal ROI Analysis. We examined prefrontal activation using separate ROI analyses for each of the regions of the prefrontal cortex that were identified in whole-brain analyses of Regulation (Decrease Negative > Look Negative). Functional ROIs were created by intersecting a structural mask of frontal regions from the Harvard-Oxford subcortical atlas (50% threshold) with the map of activation in the entire sample during reappraisal, resulting in 9 prefrontal ROIs listed in Supplemental Table S2, including the superior frontal gyrus, middle frontal gyrus, inferior frontal gyrus, frontal pole, and dorsal anterior cingulate cortex. Linear regression models were computed with depression symptoms at baseline and follow up as the outcome measure, prefrontal activation during regulation, history of trauma and their interactions as predictors, while controlling for socioeconomic status, non-white status, sex, and age. All predictors were mean-centered. When examining follow up depression as an outcome of interest, days between baseline and follow up and baseline symptoms of depression were also included as covariates. Analyses were repeated for anxiety symptoms.

Daily Emotion Regulation Strategy. To evaluate whether daily use of reappraisal moderated the association of maltreatment with depression, we estimated a linear regression model with baseline and follow-up depression as the outcome, reappraisal tendency, abuse severity, and their interactions as predictors, controlling for socioeconomic status, non-white status, sex, and age. When examining follow up depression, days between baseline and follow up and baseline symptoms of depression were also included as covariates. Analyses were repeated for anxiety symptoms.

ACCEPTED MANUSCRIPT

Supplemental Results

Behavioral Task Effects

When examining task performance across the entire group, participants rated negative pictures (M=1.62, SD=0.64) as more emotionally evocative than neutral pictures (M=0.18, SD=0.36; $t(150)=28.1$, $p<.001$), as expected. In addition, subjective ratings indicated that participants had significantly lower affective responses to negative pictures following the appraisal conditions (M=1.24, SD=0.68) as compared to passively looking at negative pictures ($t(150)=10.86$, $p<.001$), giving confidence that subjects performed the task successfully. When examining whether subjective reports of reappraisal moderates the relationship between maltreatment and depression, effects are specific to amygdala modulation and are not associated with subjective ratings of affect, which is consistent with prior work from our lab and others (12, 22). More specifically, the relationship between maltreatment and risk for depression is not moderated by reductions in subjective affect ratings during reappraisal (interaction term: $b=-0.928$, $t=-0.316$, $p=.753$).

Whole-brain Task Effects

Across the entire sample, the use of cognitive reappraisal elicited the expected pattern of activation of frontoparietal regions and decreased activation of the amygdala when compared to passive viewing of negative emotional stimuli (Supplemental Figure S1).

ACCEPTED MANUSCRIPT

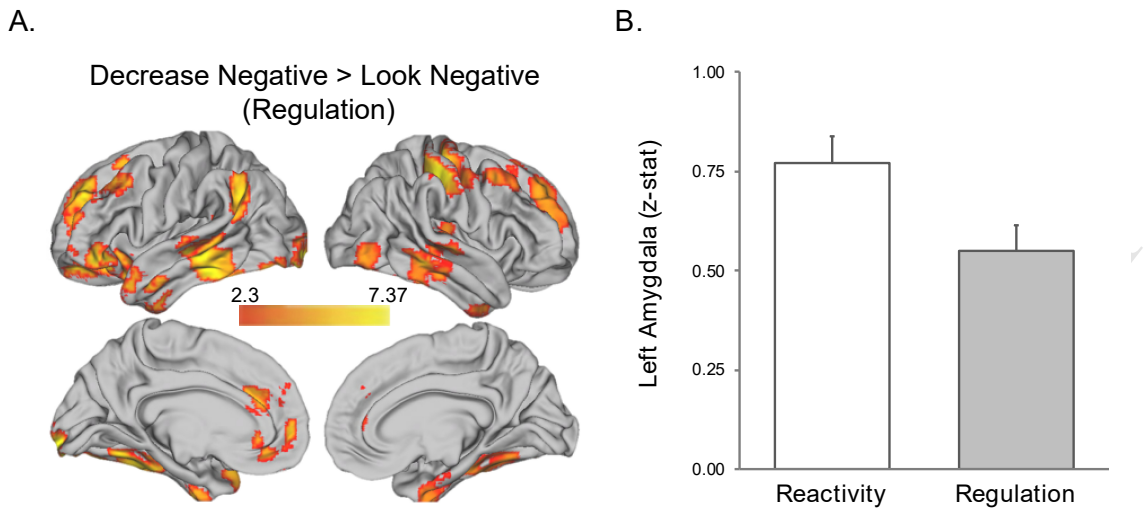


Figure S1. Neural activity during cognitive reappraisal. A. Regions that show increased functional recruitment during cognitive reappraisal relative to passive viewing of negative emotional stimuli (decrease negative > look negative). Note: Cluster-level correction applied in FSL, $z > 2.3$ was the primary threshold, and $p < .01$ was the cluster-level threshold. Color bar indicates magnitude of z-statistic B. Extracted z-stats from left amygdala ROI shows modulation of activation during passive viewing (reactivity; white bar) and reappraisal (regulation; gray bar) of negative compared to neutral images for illustrative purposes. Error bars indicate standard error of the mean.

Group Differences

Findings indicated no group differences in the extent of amygdala modulation during reappraisal (t(149)=-0.163, p=.871), recruitment of prefrontal ROIs (rSFG: t(149)=1.359, p=.176, rdACC: t(149)=0.600, p=.549), or tendency to engage in reappraisal day-to-day (t(149)=-0.118, p=.906). However, compared to those with no history of maltreatment, youth with history of maltreatment reported less reductions in subjective affect when engaging in reappraisal compared to passive viewing of negative pictures (history of maltreatment: M=-0.244, SD=0.391; no history: M=-0.524, SD=0.420; independent t-test: t(149)=4.241, p<.001).

ACCEPTED MANUSCRIPT

Primary ROI Resilience Analysis

When examining the relationship between emotion regulation efficacy and risk for future depression, results revealed a that participants' ability to modulate amygdala reactivity during reappraisal moderated the relationship between maltreatment history and depression scores at follow up. After controlling for sex, age, number of days between baseline and follow up visits, and baseline CDI depression score, CDI depression score at follow up was significantly associated with the interaction between left amygdala activation during reappraisal (Decrease Negative > Look Negative) and maltreatment history (b=4.20, t=2.41, p=.018). Aside from baseline depression (b=0.670, t=7.560, p<.001), no other covariates significantly predicted follow-up depression (ps>.05). This pattern of resilience was not evident for anxiety symptoms or baseline depression (ps>.05). Findings suggest that greater reappraisal ability, as indexed by successful amygdala modulation, may buffer maltreated youth from negative effects of adversity and lower risk for future depression.

Exploratory ROI Analysis

We sought to explore whether engagement of prefrontal regions during reappraisal moderated the relationship between maltreatment and psychopathology. Findings revealed that activation of the right superior frontal gyrus (SFG) and right dorsal anterior cingulate cortex (dACC) moderated the relationship between history of maltreatment and current depression (activation-by-maltreatment interaction SFG: b=-3.11, t=-2.50, p=.014; dACC: b=-2.65, t=-1.95, p=.052). Aside from maltreatment history (b=7.025, t=4.825, p<.001) and age (b=0.704, t=3.349, p=.001), no other covariates significantly predicted current depression (ps>.05). This interaction was not

ACCEPTED MANUSCRIPT

evident for other exploratory frontal ROI regions, anxiety symptoms, or follow up depression (*p*s>.05).

Daily Emotion Regulation Strategy

When examining daily tendency to use emotion regulation strategies, findings showed that reappraisal tendency significantly moderated the relationship between abuse severity and baseline depression (*b*=-0.044, *t*=-2.420, *p*=.017). The main effect of maltreatment severity (*b*=2.326, *t*=4.522, *p*<.001) and age (*b*=0.589, *t*=3.094, *p*=.002) were also significantly related to baseline depression. This pattern was not observed for anxiety symptoms or follow up depression (*p*s>.05).

Effects of Age and Sex

When examining the impact of age on reappraisal, our findings indicated that age has no significant effect on brain function during cognitive reappraisal in the ROIs implicated in resilience. Specifically, the main effect of age on amygdala modulation during reappraisal is not significant (*b*=0.017, *t*=0.802, *p*=.424) and neither is the effect of age on recruitment of prefrontal regions during reappraisal (right SFG: *b*=-0.032, *t*=-1.171, *p*=.243; right dACC: *b*=-0.023, *t*=-0.933, *p*=.352). While age was not related to the reported tendency to engage in daily reappraisal (*b*=0.257, *t*=1.124, *p*=.263), age was significantly associated with subjective ratings of affect modulation during the reappraisal task (*b*=-0.027, *t*=-2.078, *p*=.040), wherein older participants reported greater reductions in negative affect during regulation.

To examine sex effects, we computed independent t-tests for each ROI implicated in resilience with sex as the between-subjects factor. While there was no significant main effect of sex on amygdala modulation or recruitment of the right dACC (*p*s>.499), recruitment of the right

ACCEPTED MANUSCRIPT

SFG differed significantly between males and females (t(149)=-2.048, *p*=0.042), wherein males recruited to a greater degree than females (males: M=0.342, SD=0.872; females: M=0.052, SD=0.865). All analyses reported in the manuscript control for age and sex, in addition to socioeconomic status and race, as nuisance regressors.

Cross-measure Correlations

Trait-level tendency to use reappraisal (ERQ) is not correlated with subjective reports of reappraisal success, amygdala modulation or right SFG recruitment during reappraisal (*p*s>.05). It is, however, significantly correlated with recruitment of the right dACC during reappraisal (r(149)=.17, *p*=0.043). Amygdala modulation during reappraisal is correlated with recruitment of the right SFG (r(149)=.18, *p*=.029) and dACC (r(149)=.18, *p*=.020; Supplemental Table S3).

Sensitivity Analyses

Correction for Multiple Comparisons. P-values were corrected using the *p.adjust* package in R by controlling for the false discovery rate using the Benjamini–Hochberg method (23). Our primary analysis examining whether children with history of maltreatment, in particular, are protected from risk for depression by successful down-modulation of amygdala during reappraisal was corrected for 2 comparisons (right and left amygdala) and the significant interaction in the left amygdala survived correction for multiple comparisons (*p*=.035). Exploratory analyses in the prefrontal ROI regions were corrected for 9 comparisons (see Supplemental Table S2 for all prefrontal ROIs) and initial findings did not survive correction (*p*s>.123).

Controlling for Medication Use. Any participants taking short-acting medications, such as ADHD medications, abstained from taking medication on the day of the scan. Nonetheless, some

ACCEPTED MANUSCRIPT

participants were on other psychotropic medications that could not be temporarily discontinued, and we have conducted all key analyses controlling for medication status (i.e., taking medication on day of scan). When recomputing the interaction effect of amygdala modulation and history of maltreatment on follow up depression, findings remain significant (b=4.25, t=2.450, p=.016) when controlling for medication use. Similarly, findings examining prefrontal recruitment during reappraisal also remain significant (rSFG: b=-3.047, t=-2.490, p=.014; rdACC: b=-2.560, t=-1.974, p=.050).

Pubertal Effects. We have also conducted sensitivity analyses that include pubertal status as measured by the Tanner scale instead of age as a covariate. Key findings remained largely the same. However, the finding that activation in the prefrontal ROI right dACC interacts with maltreatment history to predict baseline depression changed from significant to trending towards significance (p=.08) in the same direction and pattern. Thus, we conclude that these findings are generally robust.

Controlling for IQ. Finally, we conducted all analyses controlling for IQ. Key findings remained largely the same, though the finding that activation in the prefrontal ROI right dACC interacts with maltreatment history to predict baseline depression changed from significant to marginally significant (p=.06) in the same direction and general pattern. Thus, findings are generally robust to controlling for IQ.

rdACC ROI Sensitivity Analyses. The right dACC ROI contains a cluster that is fairly posterior to other clusters in that ROI. However, according to the Harvard-Oxford Atlas, this cluster resides in the ACC. Nonetheless, we have computed the right dACC findings without this cluster as a sensitivity analysis and results indicate that findings hold at a marginally significant level of p=.06 in the same pattern of findings.

ACCEPTED MANUSCRIPT

Supplemental Tables

Table S1. Task-related neural activation across the whole sample

Emotion Regulation Condition (Decrease Negative > Look Negative)						
Anatomical Region	*x*	*y*	*z*	*voxels*	*z-max*	*p-value*
Right postcentral gyrus	42	-26	50	5354	7.37	<.0001
Right precentral gyrus						
Left occipital pole	-18	-100	-10	6931	6.18	<.0001
Left temporal occipital FG						
Left middle temporal gyrus						
Left inferior temporal gyrus						
Left cerebellum						
Left superior temporal gyrus						
Right middle temporal gyrus	64	-36	-10	3937	4.93	<.0001
Right temporal occipital FG						
Right temporal FG						
Right inferior temporal gyrus						
Right superior temporal gyrus						
Left frontal pole, left IFG, left insula	-38	38	-12	5438	4.75	<.0001
Left putamen						
Right frontal pole	46	36	-12	1295	4.17	0.006
Right MFG						
Right SFG						
Left SFG	-20	20	56		3.2	
Left MFG	-32	20	48		2.97	

Note. ACC=anterior cingulate cortex; IFG=inferior frontal gyrus; MFG=middle frontal gyrus; SFG=superior frontal gyrus; FG=fusiform gyrus; PHG=parahippocampal gyrus.

ACCEPTED MANUSCRIPT

Table S2. List of ROIs and size

ROI Masks	*voxels (k)*
Left Amygdala	485
Right Amygdala	18
Left Superior Frontal Gyrus	7
Right Superior Frontal Gyrus	76
Left Frontal Pole	1076
Right Frontal Pole	1206
Left Dorsal Anterior Cigulate	47
Right Dorsal Anterior Cingulate	30
Left Inferior Frontal Gyrus	95
Left Middle Frontal Gyrus	208
Right Middle Frontal Gyrus	206

Table S3. Cross-Measure Correlations

Cross-Measure Correlations	Subjective Report	Amygdala Modulation	rSFG Recruitment	rdACC Recruitment	Reappraisal Tendency
Subjective Report of Reappraisal					
pearson r	1	0.08	0.01	0.00	-0.09
p-value	0	0.358	0.919	0.974	0.298
Amygdala Modulation during Reappraisal					
pearson r	0.08	1	0.18	0.18	0.01
p-value	0.358	0	0.029*	0.020*	0.873
rSFG Recruitment during Reappraisal					
pearson r	0.01	0.18	1	0.15	0.03
p-value	0.919	0.029*	0	0.071	0.707
rdACC Recruitment during Reappraisal					
pearson r	0.00	0.18	0.15	1	0.17
p-value	0.974	0.020*	0.071	0	0.043*
Tendency to Engage in Reappraisal in Daily Life					
pearson r	-0.09	0.01	0.03	0.17	1
p-value	0.298	0.873	0.707	0.043*	0

ACCEPTED MANUSCRIPT

Supplemental References

1. Bifulco A, Brown GW, Harris TO (1994): Childhood Experience of Care and Abuse (CECA): a retrospective interview measure. *Journal of Child Psychology and Psychiatry*. 35: 1419–1435.

2. Steinberg AM, Brymer MJ, Kim S, Briggs EC, Ippen CG, Ostrowski SA, *et al.* (2013): Psychometric properties of the UCLA PTSD reaction index: part I. *Journal of traumatic stress*. 26: 1–9.

3. Raviv A, Erel O, Fox NA, Leavitt LA, Raviv A, Dar I, *et al.* (2001): Individual measurement of exposure to everyday violence among elementary schoolchildren across various settings. *Journal of Community Psychology*. 29: 117–140.

4. Bernstein DP, Ahluvalia T, Pogge D, Handelsman L (1997): Validity of the Childhood Trauma Questionnaire in an adolescent psychiatric population. *Journal of the American Academy of Child & Adolescent Psychiatry*. 36: 340–348.

5. Walker EA, Unutzer J, Rutter C, Gelfand A, Saunders K, VonKorff M, *et al.* (1999): Costs of Health Care Use by Women HMO Members With a History of Childhood Abuse and Neglect. *Archives of General Psychiatry*. 56: 609.

6. Finkelhor D, Hamby SL, Ormrod R, Turner H (2005): The Juvenile Victimization Questionnaire: reliability, validity, and national norms. *Child abuse & neglect*. 29: 383–412.

7. Bae Y (2012): Test Review: Children's Depression Inventory 2 (CDI 2). *Journal of Psychoeducational Assessment*. 30: 304–308.

8. Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, Neer SM (1997): The screen for child anxiety related emotional disorders (SCARED): Scale construction and psychometric characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*. 36: 545–553.

9. Gross JJ, John OP (2003): Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*. 85: 348–362.

10. Ochsner KN, Gross JJ (2005): The cognitive control of emotion. *Trends in Cognitive Sciences*. 9: 242–249.

11. McRae K, Gross JJ, Weber J, Robertson ER, Sokol-Hessner P, Ray RD, *et al.* (2012): The development of emotion regulation: an fMRI study of cognitive reappraisal in children, adolescents and young adults. *Social cognitive and affective neuroscience*. 7: 11–22.

12. Silvers JA, Insel C, Powers A, Franz P, Helion C, Martin R, *et al.* (2017): The transition from childhood to adolescence is marked by a general decrease in amygdala reactivity and an

ACCEPTED MANUSCRIPT

affect-specific ventral-to-dorsal shift in medial prefrontal recruitment. *Developmental Cognitive Neuroscience*, Sensitive periods across development. 25: 128–137.

13. Cox RW (1996): AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research*. 29: 162–173.

14. Jenkinson M, Bannister P, Brady M, Smith S (2002): Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*. 17: 825–841.

15. Smith SM, Brady JM (1997): SUSAN—A New Approach to Low Level Image Processing. *International Journal of Computer Vision*. 23: 45–78.

16. Behzadi Y, Restom K, Liau J, Liu TT (2007): A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI. *Neuroimage*. 37: 90–101.

17. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC (2011): A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*. 54: 2033–2044.

18. McLaughlin KA, Peverill M, Gold AL, Alves S, Sheridan MA (2015): Child maltreatment and neural systems underlying emotion regulation. *Journal of the American Academy of Child & Adolescent Psychiatry*. 54: 753–762.

19. Silvers JA, McRae K, Gabrieli JD, Gross JJ, Remy KA, Ochsner KN (2012): Age-related differences in emotional reactivity, regulation, and rejection sensitivity in adolescence. *Emotion*. 12: 1235.

20. Ochsner KN, Bunge SA, Gross JJ, Gabrieli JD (2002): Rethinking feelings: an FMRI study of the cognitive regulation of emotion. *Journal of cognitive neuroscience*. 14: 1215–1229.

21. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012): Fsl. *Neuroimage*. 62: 782–790.

22. McLaughlin KA, Peverill M, Gold AL, Alves S, Sheridan MA (2015): Child Maltreatment and Neural Systems Underlying Emotion Regulation. *J Am Acad Child Adolesc Psychiatry*. 54: 753–762.

23. Benjamini Y, Hochberg Y (1995): Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 57: 289–300.

Maternal mediation, stress inoculation, and the development of neuroendocrine stress resistance in primates

Karen J. Parker*, Christine L. Buckmaster, Karan Sundlass, Alan F. Schatzberg, and David M. Lyons

Department of Psychiatry and Behavioral Sciences, Stanford University, 1201 Welch Road, MSLS P104, Stanford, CA 94305-5485

Edited by Bruce S. McEwen, The Rockefeller University, New York, NY, and approved January 3, 2006 (received for review August 1, 2005)

The stress inoculation hypothesis presupposes that brief intermittent stress exposure early in life induces the development of subsequent stress resistance in human and nonhuman primates. Rodent studies, however, suggest a role for maternal care rather than stress exposure per se (i.e., the maternal mediation hypothesis). To investigate these two hypotheses, we examined maternal care and the development of stress resistance after exposure to brief intermittent infant stress (IS), mother–infant stress (MIS), or no stress (NS) protocols administered to 30 monkeys between postnatal weeks 17 and 27. Unlike rodents, the IS condition did not permanently increase primate maternal care, nor did measures of total maternal care predict subsequent offspring hypothalamic–pituitary–adrenal-axis responsivity. Although MIS infants received less maternal care than IS and NS infants, both IS and MIS monkeys developed subsequent stress resistance. These findings indicate that rearing differences in the development of stress resistance are more closely related to differences in prior stress exposure than to differences in maternal care. A second experiment confirmed this conclusion in a different cohort of 25 monkeys exposed as infants to high foraging-demand (HFD) or low foraging-demand (LFD) conditions. HFD infants exhibited intermittent elevations in cortisol levels and received less maternal care than LFD infants. In keeping with a key prediction of the stress inoculation hypothesis, HFD males responded to stress in adulthood with diminished hypothalamic–pituitary–adrenal-axis activation compared with LFD males. Results from both experiments demonstrate that stress inoculation, rather than high levels of maternal care, promotes the development of primate stress resistance.

hypothalamic–pituitary–adrenal axis | monkey | maternal care | resilience

People with the capacity to maintain healthy emotional functioning in the aftermath of stressful experiences are said to be resilient, or stress resistant (1, 2). Researchers have sought to identify attributes associated with stress resistant individuals, with the expectation that understanding the etiology of stress resistance may lead to the prevention of stress-related psychiatric disorders. One intriguing finding to emerge from this retrospective research has been that stress resistance is associated with childhood exposure to mildly stressful events (1, 3). Prospective longitudinal studies of nonhuman primates support this finding, because monkeys exposed to brief, 1-h periods of maternal-separation stress exhibit diminished anxiety and attenuated hypothalamic–pituitary–adrenal (HPA)-axis responses to subsequent stressors compared with unmanipulated monkeys (4).

Little is known about the psychobiological mechanisms that promote the development of stress resistance in human or nonhuman primates, but studies of rodents suggest a possible role for maternal care. Similar to monkeys, neonatal rodents exposed to brief repeated maternal separations of 3–15 min in duration (i.e., postnatal handling) exhibit diminished emotionality and attenuated HPA-axis responses to subsequent stressors (5, 6). The development of rodent stress resistance is thought to be maternally mediated (7), because brief intermittent separations stimulate increased maternal licking and grooming not only at reunion but across pup development (8). These permanent changes in maternal behavior increase the expression of glucocorticoid receptors in the hippocampus, thereby enhancing sensitivity to glucocorticoid-feedback inhibition during stress exposure later in life (5). Importantly, the emotional and neuroendocrine stress resistance observed in briefly separated rodents can be replicated in unmanipulated offspring that naturally receive high levels of maternal care during development (9). That increased maternal care is sufficient to produce stress resistance in rodents is corroborated by evidence that the amount of maternal stimulation received by pups during development is negatively correlated with later stress-induced HPA-axis activation (9, 10).

The utility of rodent research in modeling the etiology of human stress resistance is limited, however, by critical developmental differences in rodent and primate HPA-axis physiology. The brief maternal separation paradigm coincides with the rodent stress-hyporesponsive period (SHRP), during which time mice and rats do not respond, or respond only weakly, to mild stressors with HPA-axis activation (11, 12). High levels of maternal stimulation maintain the pup's SHRP status, with removal of maternal regulation resulting in gradual disinhibition of HPA-axis activity. Diminished maternal stimulation that fails to adequately inhibit activation of the pup's HPA-axis during the SHRP, as documented for longer maternal separations of 3 h in duration (13), promotes stress hyperresponsivity later in life. Although it has not been systematically studied, human and nonhuman primate infants do not appear to exhibit a SHRP and are capable of exhibiting stress responses throughout lifespan development (14–17). Primate infants, unlike rodent pups, respond to brief intermittent maternal separations with immediate HPA-axis activation (18) and develop subsequent neuroendocrine stress resistance even after maternal separations of 4–6 h in duration (19). If the same psychobiological mechanisms "program" the stress response in rodents and primates, HPA-axis activation during primate development should lead to later stress hyperresponsivity. Yet, in primates, a paradigm that induces anxiety and activates the HPA-axis leads, instead, to emotional and neuroendocrine stress resistance (4, 19).

An alternative to the maternal mediation hypothesis is the stress inoculation hypothesis, which is based on the notion that mild stress, and the acute anxiety and HPA-axis activation that it engenders, is necessary for the development of subsequent stress resistance. Several theorists have likened the development of stress resistance to acquired immunity (1, 3). Immunity, whether naturally occurring or therapeutically induced by inoc-

---

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: ACTH, adrenocorticotropic hormone; HFD, high foraging-demand; HPA, hypothalamic–pituitary–adrenal; IS, infant stress; LFD, low foraging-demand; MIS, mother–infant stress; NS, no stress; SHRP, stress-hyporesponsive period.

*To whom correspondence should be addressed. E-mail: kjparker@stanford.edu.

© 2006 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0506571103

ulation, derives from exposure to a mild version of a pathogen that strengthens immunological resistance. Similarly, early exposure to mild stress may "inoculate" the developing organism, thereby enhancing resistance to subsequent stressors. In the following experiments, we investigated the roles of maternal mediation and stress inoculation in the development of stress resistance in squirrel monkeys.

Experiment 1

The goals of this experiment were twofold: (*i*) to determine whether brief intermittent separations permanently increase primate maternal behavior during development, and (*ii*) to test whether maternal care received during development predicts subsequent offspring HPA-axis responsivity. Thirty infants from three rearing conditions were studied. In the first condition, only infants were exposed to brief intermittent separations from the natal group for 10 weekly 1-h sessions [infant stress (IS)]. In the second condition, mothers and infants were removed together from the natal group and exposed to an otherwise identical stress protocol [mother–infant stress (MIS)]. Both IS and MIS protocols acutely induce anxiety and activate the HPA-axis in young monkeys (18, 20). In the third condition, mothers and infants remained undisturbed in their natal groups [no stress (NS)]. Maternal care was assessed every other week during three distinct periods in the home cage (i.e., before stress exposure, immediately after stress exposure, and 24 h after stress exposure). Eleven weeks after completion of the rearing protocols, a 30-min novelty stress test was administered to examine rearing differences in offspring HPA-axis responsivity.

Research has shown that IS infants develop neuroendocrine stress resistance, whereas NS infants do not (19). The maternal-mediation hypothesis therefore predicts that IS infants will consistently receive more maternal care than NS infants. Although it is unknown whether MIS infants develop stress resistance, this condition was included because maternal stress diminishes primate maternal care (21). If MIS infants receive less or equal amounts of maternal care compared with control infants, the maternal mediation hypothesis predicts that MIS infants will not develop neuroendocrine stress resistance. The stress inoculation hypothesis, in contrast, predicts the development of neuroendocrine stress resistance in all offspring exposed to brief intermittent stress protocols, because these protocols acutely induce anxiety and activate the HPA-axis.

Results. ***Rearing differences in the frequency and duration of maternal behavior.*** Rearing protocols did not permanently alter the frequency or duration of any maternal behavior measure across development, because no rearing differences were evident in the periods immediately before or 24 h after stress exposure. Several transient effects were observed immediately after stress exposure, however, including a main effect of rearing condition on the duration of dorsal contact (i.e., the species-typical infant riding posture on the mother's shoulders and upper back) ($F_{2,26} = 5.304$, $P = 0.012$). Upon returning to the home cage, IS infants spent more time clinging to their mothers than did NS infants (39 ± 3 vs. 19 ± 6 min; $P = 0.011$). IS infants also clung to their mothers more than did MIS infants (MIS = 26 ± 4 min), but this difference failed to reach statistical significance. MIS and NS infants did not differ from one another. A significant protocol-week by rearing-condition interaction was found for nursing duration immediately after stress exposure ($F_{8,104} = 4.37$, $P = 0.004$). However, post hoc tests at each week revealed no consistent rearing differences, because both IS and MIS infants differed from NS infants (and not each other) at only two of the five time points (i.e., weeks 3 and 5). These data indicate that IS infants generally received more maternal care than NS infants immediately after stress exposure, whereas MIS infants did not significantly differ from either group.

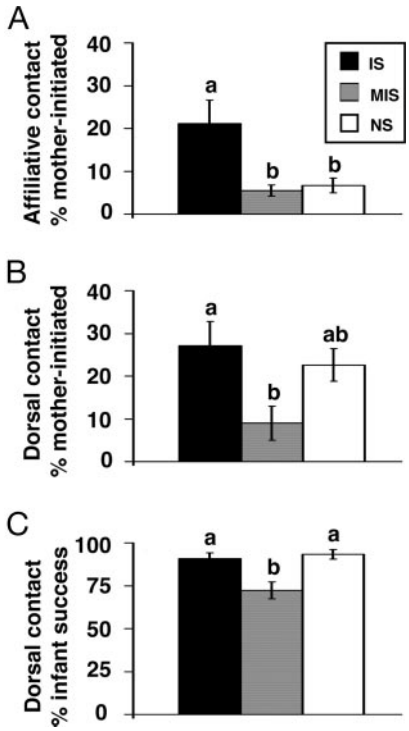


Fig. 1. Rearing differences in mother–infant social transactions. Affiliative (*A*) and dorsal (*B*) contact attempts initiated by mothers relative to infants and total infant-initiated dorsal contact attempts accommodated by mothers (*C*) are presented as mean ± SEM for monkeys previously exposed to IS ($n = 11$), MIS ($n = 10$), or NS ($n = 9$) rearing conditions. Groups without shared letters differ significantly ($P < 0.05$), whereas groups with shared letters do not.

Rearing differences in mother–infant social transactions. Enduring rearing differences in mother–infant transactions were observed across all three observation period types (Fig. 1). A significant effect of rearing condition was found for the percentage of total affiliative contact attempts initiated by mothers relative to their infants ($F_{2,26} = 4.964$, $P = 0.015$). IS mothers made more affiliative contact attempts (21%) than either MIS (6%; $P = 0.02$) or NS (7%; $P = 0.049$) mothers, whereas MIS and NS mothers did not differ. Similarly, rearing differences were found for the percentage of total dorsal contact attempts initiated by mothers relative to infants ($F_{2,26} = 3.787$, $P = 0.036$). IS mothers initiated dorsal contact with infants 27% of the time, whereas MIS mothers initiated dorsal contact only 9% of the time ($P = 0.043$). Although NS mothers' dorsal contact initiatives more closely resembled those of IS mothers (23%), they did not significantly differ from either group. Finally, a rearing difference in the percentage of total infant-initiated dorsal contact attempts that were accommodated by mothers was found ($F_{2,26} = 8.340$, $P = 0.002$). MIS infants' attempts to attain dorsal contact were accommodated 72% of the time, whereas IS and NS infants dorsal contact attempts were, respectively, accommodated 91% ($P = 0.009$) and 93% ($P = 0.003$) of the time. These results indicate that, across observation periods, MIS infants received diminished maternal care relative to IS and, at times, NS infants.

Neuroendocrine stress responsivity and maternal care. Rearing differences were found for baseline cortisol levels ($F_{2,26} = 10.739$, $P < 0.0001$), such that both IS ($P = 0.048$) and MIS ($P < 0.0001$) offspring exhibited diminished baseline cortisol levels compared with NS offspring. To control for these differences, baseline cortisol levels were used as a statistical covariate to analyze rearing differences in poststress cortisol levels. Consistent with

PSYCHOLOGY

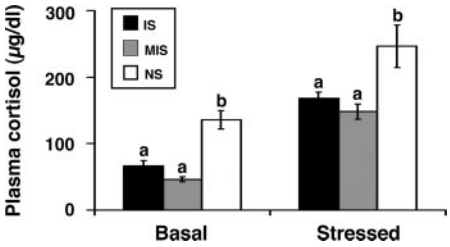


Fig. 2. Plasma cortisol levels at baseline and after stress exposure are presented as mean ± SEM for juvenile monkeys previously exposed to IS (n = 11), MIS (n = 10), or NS (n = 9) rearing conditions. Groups without shared letters (a, b) differ significantly ($P < 0.05$).

the stress inoculation hypothesis, rearing differences in stress resistance were observed ($F_{2,25} = 5.124$, $P = 0.014$). Specifically, previously stress-exposed IS ($P = 0.033$) and MIS ($P = 0.013$) offspring exhibited diminished cortisol responses to the novelty stress test compared with NS offspring (Fig. 2).

Total maternal care did not predict subsequent offspring cortisol levels at baseline or after novelty stress exposure. Although enduring rearing differences in several social-transaction measures were found, as was a transient increase in dorsal contact immediately after stress exposure, none of these measures predicted subsequent baseline- or stress-levels of cortisol. Nursing duration in the period immediately after stress exposure, however, did predict cortisol levels at baseline [$\beta = -0.525$ (df = 28); $P = 0.003$] and after novelty stress exposure [$\beta = -0.397$ (df = 28), $P = 0.030$]. To test the hypothesis that nursing duration mediates the effects of rearing condition on baseline- and stress-levels of cortisol, nursing duration and rearing condition were examined together as independent variables in the same regression equation (22, 23). Rearing-related differences in baseline ($F_{2,26} = 4.79$, $P = 0.017$) and stress ($F_{2,26} = 3.86$, $P = 0.034$) levels of cortisol remained significant, whereas nursing duration no longer predicted either cortisol measure. These results do not support a role for nursing as a mediator of the effect of rearing condition on cortisol levels.

Experiment 2

Findings from experiment 1 suggest that rearing differences in neuroendocrine stress resistance are more closely related to differences in prior stress exposure than to differences in maternal care. The primary goal of experiment 2 was to assess the generality of these findings by testing the maternal mediation and stress inoculation hypotheses in 25 monkeys exposed to a different type of mild early life stress. The protocols used in this experiment consisted of high foraging-demand (HFD) and low foraging-demand (LFD) conditions that differed in the effort required by mothers to obtain food. These daily foraging protocols did not alter infant growth, weight, nursing behavior, or solid-food consumption, but HFD infants exhibited intermittently elevated cortisol levels and received diminished maternal care compared with LFD infants across development (24, 25).

At 8 years of age, monkeys were restrained in primate chairs for two 30-min sessions that were administered 7 days apart. An intramuscular injection of saline was given 60 min before the first restraint test to examine rearing differences in HPA-axis responsivity and poststress recovery. Because neuroendocrine stress resistance in rats arises from enhanced glucocorticoid-feedback regulation of the HPA-axis stress response (5), we also tested in this experiment whether the same was true for monkeys. Thus, exogenous cortisol (i.e., 2.5 mg/kg hydrocortisone sodium succinate) was intramuscularly injected 60 min before the second restraint test. This dose of hydrocortisone inhibits stress-induced increases in squirrel monkey adrenocorticotropic hormone

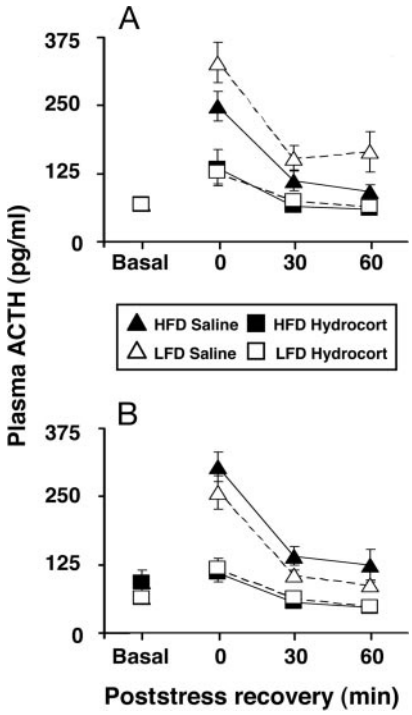


Fig. 3. Plasma ACTH levels after pretreatment with saline or hydrocortisone 0, 30, and 60 min after stress exposure are presented as mean ± SEM for male (*A*) and female (*B*) monkeys previously exposed to HFD (n = 6 males, 6 females) or LFD (n = 7 males, 6 females) rearing conditions.

(ACTH) (26) and was used to assess rearing differences in glucocorticoid-feedback inhibition of the HPA-axis stress response.

The maternal mediation and stress inoculation hypotheses make different predictions about HPA-axis responsivity in HFD and LFD monkeys. The maternal mediation hypothesis predicts that the diminished maternal behavior and elevated cortisol levels that characterize the HFD condition will promote increased stress responsivity in HFD compared with LFD monkeys. Conversely, the stress inoculation hypothesis predicts that the mild stress and elevated cortisol levels that typify the HFD condition will produce stress resistance in HFD but not LFD monkeys.

Baseline ACTH and cortisol levels did not vary across repeated samples, nor did either differ by rearing condition. Although no effect of gender on baseline ACTH levels was found, gender did influence baseline cortisol levels ($F_{1,21} = 14.785$, $P = 0.001$). Specifically, females had baseline cortisol levels that were 2-fold higher than males (186 ± 24 vs. 87 ± 9).

A significant rearing-condition by gender interaction across the saline-pretreatment poststress sample time points for ACTH was found ($F_{1,20} = 4.476$, $P = 0.047$). Further analysis revealed a main effect of rearing condition on stress responsivity for male, but not female, monkeys (Fig. 3). Although ACTH levels decreased significantly over time in all monkeys ($F_{2,40} = 22.461$, $P < 0.0001$), only adult males previously exposed to the HFD condition responded to restraint stress with diminished ACTH levels compared with males from the LFD condition ($F_{1,10} = 6.253$, $P = 0.031$). This rearing effect was also apparent for time-integrated poststress measures of ACTH in male ($F_{1,10} = 5.851$, $P = 0.036$), but not female, monkeys.

Unlike ACTH, cortisol levels after the saline pretreatment continued to rise significantly across the poststress measurement period. In keeping with the gender differences in baseline cortisol levels described above, time-integrated poststress mea-

sures of cortisol in females (306 ± 25 μg/dl) were 67% greater than in males (184 ± 14 μg/dl; $F_{1,21} = 18.358$, $P < 0.0001$). However, the magnitude of poststress increases in cortisol levels above baseline did not differ by gender. Cortisol levels at 0, 30, and 60 min after stress were, respectively, 68 ± 8, 100 ± 10, and 118 ± 8 μg/dl higher than baseline for males, and 79 ± 26, 126 ± 23, and 148 ± 26 μg/dl higher than baseline for females. No significant gender or rearing effects on poststress cortisol levels were discerned.

As expected, the hydrocortisone pretreatment robustly suppressed stress levels of ACTH relative to the saline pretreatment (Fig. 3). Hydrocortisone suppressed time-integrated poststress levels of ACTH in both males ($F_{1,11} = 99.83$, $P < 0.0001$) and females ($F_{1,10} = 69.47$, $P < 0.0001$). Within 30 min of stressor termination, ACTH levels were equal to or less than baseline in all monkeys, and glucocorticoid-feedback inhibition did not differ significantly by rearing condition.

Discussion

Findings from experiment 1 indicate that, unlike rodents, brief maternal separations do not permanently increase primate maternal behavior, nor does total maternal care predict later offspring HPA-axis stress responsivity. Although young MIS infants receive less maternal care across development than IS or NS infants, both IS and MIS monkeys develop stress resistance. Despite the fact that nursing duration in the period immediately after stress exposure predicted later cortisol measures, subsequent analyses did not support a role for nursing as a mediator of the effect of rearing condition on cortisol levels. These findings suggest that rearing differences in the development of stress resistance are more closely related to differences in prior stress exposure than to differences in maternal care. Experiment 2 confirmed these findings in a second cohort of monkeys exposed to a different type of mild early life stress. HFD infants exhibit modestly elevated cortisol levels and receive less maternal care than LFD infants during development. As adults, HFD males subsequently demonstrate diminished stress-induced HPA-axis activation compared with LFD males. Results from these two experiments suggest that stress inoculation, rather than high levels of maternal care, promotes the development of primate stress resistance.

Our findings from monkeys parallel data from rabbits that indicate that brief separations from the natal nest, independent of maternal care, are sufficient to produce stress resistance. Mother rabbits provide minimal maternal care, however, and stress resistance develops in briefly separated rabbit pups whether they are raised with or without their mothers (27). In contrast, stress resistance does not develop in motherless monkeys (28) or orphaned children (29). Combined with findings from rodent studies reviewed above, this collective evidence suggests that species differences in parental behavior (i.e., minimal vs. extensive) and HPA-axis development (i.e., presence vs. absence of the SHRP) may determine how stress resistance is programmed on a species-by-species basis.

The notion that high levels of maternal care are not required to produce stress resistance does not negate the critical role of parenting in primate development nor does it indicate that aspects of lifelong health are impervious to parental care. Indeed, mother–infant attachment relationships were firmly established before the initiation of both of our experiments, and all infants received adequate maternal care, well within the range of what is typical for squirrel monkeys (30). In humans, socioemotional support derived from close relationships clearly promotes effective coping (31). It seems likely, then, that some threshold amount of maternal care during or after stress exposure is required to bolster the young primate's coping capacity, thereby contributing to the development of stress resistance.

The importance of maternal care notwithstanding, results from our studies implicate a role for stress inoculation in the development of primate stress resistance. Whether these findings reflect specific or general stress inoculation effects merits consideration. Specific inoculation refers to the notion that exposure to a given stressor confers subsequent resistance to the same or similar stressor. General inoculation, in contrast, refers to the notion that prior exposure to one stressor strengthens resistance to a variety of stressors (4). In experiment 1, the IS and MIS protocols and the subsequent stress resistance test had several elements in common. These elements include exposure to an unfamiliar environment and separation from the natal group. IS monkeys, however, also exhibit emotional stress resistance in their home cage in the presence of the natal group (4) and demonstrate enhanced performance on an emotionally challenging cognitive test compared with NS monkeys (32). There were, likewise, considerable differences between the HFD protocol and the restraint test used in experiment 2. Although further research is required to determine whether mild early life stress exposure also confers resistance to subsequent stressors of a physiological nature (e.g., immune challenge or insulin-induced hypoglycemia), the initial results from our experiments point to some form of general stress inoculation.

Several rearing-related discrepancies were apparent both between and within experiments. A rearing difference in baseline cortisol levels, for example, was found in experiment 1 but not experiment 2. This disparity is likely attributable to age, because rearing differences in baseline cortisol levels become less detectable over development (32) and are no longer discerned by the late-juvenile period (ref. 19; and K.J.P., K. L. Rainwater, C.L.B., A.F.S., S. E. Lindley, and D.M.L., unpublished work). It is not clear why this change occurs, nor is the function of this rearing effect understood. A second disparity was evident within experiment 2. Significant rearing differences in ACTH, but not cortisol, levels were found in adult male monkeys after restraint-stress exposure. In the hour after restraint stress termination, ACTH levels steadily decreased toward baseline, whereas cortisol levels continued to rise. Because the adrenal response to stress temporally follows that of the pituitary, it is likely that our assessment period was too short to capture rearing differences in adrenal recovery. The restraint test as used in experiment 2 thus appears more suitable for detecting rearing differences in the temporal dynamics of primate pituitary compared with adrenal stress physiology.

Several gender differences were found in experiment 2. These include 2-fold higher cortisol levels in adult females compared with males at baseline and after stress exposure and demonstration of neuroendocrine stress resistance only in adult males. Stress resistance is reliably expressed equally in males and females throughout adolescence (19), but evidence from experiment 2 indicates that gender differences in stress resistance may emerge after the peripubertal period. Higher baseline glucocorticoid levels and enhanced stress responsivity have been reported for adult female rats and humans, and ovarian steroids are thought to modulate both phenomena (33, 34). In contrast, androgens, which are more prevalent in adult males than females, inhibit HPA-axis activation (35). One might conclude that pubertal changes in specific gonadal steroid levels may contribute to gender differences in adult stress resistance. An alternative possibility is that stress-inoculated female monkeys do not "lose" their stress resistant status at puberty, but rather, ovarian hormones and elevated baseline cortisol levels mask this phenomenon at certain points in the ovarian cycle. Careful monitoring of ovarian function in cycling and noncycling adult females with respect to HPA-axis physiology is required to accurately test whether female stress resistance persists into adulthood.

Neuroendocrine stress resistance in rodents is mediated, in part, by enhanced glucocorticoid-feedback sensitivity (5), but we

failed to find a similar effect in primates. These results suggest that the neural basis of stress resistance in primates may differ from that in rodents. This possibility has important implications for guiding future research, particularly in humans, which more closely resemble monkeys than rats. It should be noted, however, that only one dose of hydrocortisone was used in our study. Although the dose was carefully selected based on previous research (26), examination of multiple hydrocortisone doses is required to more fully evaluate rearing differences in primate sensitivity to glucocorticoid negative feedback.

A final topic that warrants comment is that the skewed sex ratio in experiment 1 (23 females and 7 males) limits the generalizability of the cortisol and maternal care data to males. In a separate cohort, however, IS compared with NS males were previously shown to exhibit diminished stress-induced cortisol levels (19). We likewise documented in experiment 2 that, similar to MIS males, mildly stressed HFD males that received less maternal care than control males in infancy develop subsequent stress resistance. Aspects of experiment 1 nevertheless remain to be replicated, especially the effects of maternal care on subsequent HPA-axis stress responsivity in IS and NS male monkeys.

In conclusion, it is likely that certain forms of primate stress resistance arise from manageable exposure to mild stress that stimulates anxiety and activates the HPA-axis. This formulation presupposes that anxiety and HPA-axis activation during stress exposure play key roles in promoting the development of stress resistance. With continued investigation of this hypothesis, a comprehensive understanding of the etiology and neurobiology of stress resistance may ultimately provide a foundation for improved treatment and prevention of stress-related psychiatric disorders.

Methods

Experiment 1. ***Subjects.*** Thirty squirrel monkey infants (*Samiri sciureus*) were born at Stanford University and served as subjects. Subjects were housed in wire-mesh cages (1.8 × 1.2 × 1.8 m) in 10 natal groups composed of three to four mother–infant dyads. Group composition was determined primarily by birth date to minimize developmental differences between cohabitating infants. Monkeys were housed on a 12/12 h light/dark cycle in rooms with an ambient temperature of 26°C. Cages were cleaned daily, and monkeys had ad libitum access to water, food, and toys. A sliding door in each home cage facilitated access to a portable capture cage. Monkeys were trained to enter the capture cage to facilitate experimental manipulations. All procedures were approved by Stanford University's Administrative Panel on Laboratory Animal Care.

Rearing protocols. Subjects remained undisturbed in their natal groups through postnatal week 16, at which time natal groups were assigned to one of three conditions. The IS condition included 11 offspring (7 females and 4 males). From postnatal weeks 17–27, each infant was removed from its mother and natal group for a 1-h period once a week, placed in a cage (46 × 46 × 46 cm) adjacent to unfamiliar monkeys in a different room, and temporarily deprived of all forms of maternal and natal-group contact. After the 1-h separation period, each infant was returned to the home cage. No more than one infant from each natal group was separated on a given day, and all manipulations occurred between 1400 and 1800 h. The MIS condition included 10 offspring (8 females and 2 males). This rearing protocol was identical to the IS protocol in all ways, except that each infant and its mother were removed together for the 1-h period. In the third condition, nine offspring (eight females and one male) remained undisturbed as NS controls.

Behavioral observations. Each IS and MIS mother–infant dyad was observed during three distinct 15-min periods in the home cage: (*i*) before stress exposure, (*ii*) immediately after stress exposure, and (*iii*) 24 h after stress exposure. These observations were conducted for the first, third, fifth, seventh, and ninth stress exposures, for a total of 15 observation periods per dyad. Observations of NS mother–infant dyads occurred at matched time points to facilitate comparisons between rearing conditions.

Squirrel monkey mothers, unlike rodent mothers, do not lick or groom their infants. An established squirrel monkey ethogram was thus used to select maternal behavior measures (30). During postnatal weeks 17–27, infants are attaining physical, if not psychological, independence from their mothers. Infants at these ages locomote independently, consume solid foods, and nurse infrequently (36). Although no longer in constant physical contact, mothers still provide maternal care and engage in a variety of complex social interactions with their infants.

By using a computer-aided recording program, mother–infant behavior was collected by a trained observer. Durations of a given behavior were calculated by summing the elapsed period between initiation and termination. Bouts of a given behavior were distinct if separated by at least 3 seconds. The following measures were summarized for each observation period: (*i*) dorsal contact (infant observed in the species-typical riding posture on the mother's shoulders and upper back), (*ii*) affiliative contact (mother and infant observed in side-by-side huddling), and (*iii*) nursing contact (infant observed in the species-typical ventral nursing position). A composite measure of maternal care was created by combining dorsal, affiliative, and nursing contact scores.

Social transactions between mothers and infants were also recorded for each observation period (24). Transactions were initiated by attempts to change the immediate state of association between two individuals by means of either make-or-break dorsal contact initiatives or make-or-break affiliative contact initiatives. Successful attempts were scored whenever make-or-break contact initiatives were accommodated by the target. Failed attempts were scored when either initiative by an actor was resisted by the target.

Neuroendocrine stress responsivity and hormone quantification. At an average of 38 weeks of age, a 30-min novelty stress test was administered to examine rearing differences in offspring HPA-axis responsivity. Each mother–offspring dyad was transported to a cage (60 × 60 × 90 cm) in an unfamiliar room that did not contain other monkeys. Both the cage and room used for stress testing were different from those used for the rearing protocols. Tests occurred between 1500 and 1800 h.

Blood samples were collected from offspring 10 days before the experimental manipulations to establish cortisol measures in an undisturbed state. Blood samples were also collected immediately after testing to examine cortisol levels after novelty exposure, and a subset of these data was reported in ref. 4. All blood samples were collected between 1530 and 1800 h to control for circadian variation (37). Femoral blood samples were collected within 3 minutes of capture from manually restrained monkeys with single-use syringes containing 20 μl of ethylenediamine tetraacetic acid. Each sample was transferred to a chilled tube and centrifuged at 4°C, and the plasma fraction was stored at −80°C. Cortisol was measured in duplicate by using an established radioimmunoassay as described in ref. 37.

Data analysis. The effects of rearing condition on mother–infant-behavior measures and cortisol levels were assessed with ANOVA. Rearing condition and gender were between-subjects factors, and protocol week (1, 3, 5, 7, and 9), and observation period (before stress exposure, immediately after stress exposure, and 24 h after stress exposure) were within-subjects, repeated-measures factors. The Geisser–Greenhouse correction was used to adjust for multiple comparisons across repeated-measures factors. Tukey post hoc comparisons were used to discern significant rearing differences. Simple and multiple linear-regression models were used to investigate whether ma-

ternal behavior predicted offspring cortisol levels and to test whether rearing condition effects on cortisol levels were mediated by maternal behavior, respectively (22, 23). All test statistics were evaluated with two-tail probabilities ($P < 0.05$), and descriptive statistics are presented as mean ± SEM.

Experiment 2. ***Subjects, rearing protocol, and subsequent husbandry.*** Twenty-five squirrel monkeys born at Stanford University served as subjects. Subjects were housed under standard laboratory conditions in eight natal groups, each composed of three to four mother–infant dyads. Groups were randomized at 10 weeks of age to HFD (six male and six female infants) or LFD (seven male and six female infants) rearing conditions. The HFD condition consisted of 120% (by weight), and the LFD condition consisted of 600% (by weight) of normal daily food intake buried in foraging boards as described in detail in ref. 24. These foraging protocols were administered daily and lasted 12 weeks. All monkeys were housed under standard laboratory conditions thereafter. Mothers were removed after weaning at 36 weeks, and offspring were housed with two to three same-sex peers.

Neuroendocrine stress responsivity and hormone quantification. At an average of 8.4 years of age, adult monkeys were restrained for two separate 30-min sessions in a standard primate chair. Chair restraint allows for slight movement and does not induce pain or inflict tissue damage (26). Restraint stress was used because adult squirrel monkeys do not respond to novelty stress as robustly as do infants (38). Tests were administered during the breeding season, when females exhibit ovarian hormone cyclicity, to best approximate adult human physiology (39).

The first test was preceded 60 min earlier by an intramuscular injection of saline to examine rearing differences in HPA-axis stress responsivity. Seven days later, an intramuscular injection of exogenous cortisol (i.e., 2.5 mg/kg hydrocortisone sodium succinate; Upjohn, Kalamazoo, MI) was likewise administered 60 min before the second test to examine rearing differences in glucocorticoid-feedback inhibition of the HPA-axis stress response.

Immediately after completion of each test, blood samples were collected, and monkeys were returned to the home cage. Additional samples were then collected 30 and 60 min later to provide posttest measures of recovery. Blood samples were also collected from otherwise undisturbed monkeys in the home cage 7 days before and 7 days after experimental manipulations to assess baseline hormone levels.

Blood samples were obtained as described in experiment 1 between 1330 and 1430 h to control for circadian variation (37). Plasma levels of ACTH and cortisol were measured in duplicate by using established radioimmunoassays (37). The cortisol assay does not distinguish between endogenous and exogenous cortisol (i.e., hydrocortisone), and, therefore, ACTH levels served as the primary index of stress responsivity and glucocorticoid-feedback regulation (26).

Data analysis. Baseline hormone measures were analyzed for rearing condition, gender, and time effects by using repeated-measures ANOVA. Time-integrated poststress hormone measures were determined with the trapezoidal rule to estimate the area under each monkey's saline and hydrocortisone time course curve (26). Time course (0, 30, and 60 min after restraint) and time-integrated poststress measures were then analyzed for rearing condition, gender, and hydrocortisone-pretreatment effects with baseline hormone levels controlled as a statistical covariate. All test statistics were evaluated with two-tailed probabilities ($P < 0.05$), and descriptive statistics are presented as mean ± SEM.

We thank Dr. Seymour Levine, Dr. Shelly Flagel, and Michael Heller for their thoughtful comments on this manuscript and Dr. Helena Kraemer for statistical guidance. This work was supported by National Institute of Mental Health Grants MH66537, MH47573, and MH50605; National Institute on Drug Abuse Grant DA16902; the Nancy Pritzker Network (New York); and a Stanford University School of Medicine Dean's Fellowship.

1. Rutter, M. (1993) *J. Adolesc. Health* **14,** 626–631; 690–696.
2. Masten, A. S. (2001) *Am. Psychol.* **56,** 227–238.
3. Boyce, W. T. & Chesterman, E. (1990) *J. Dev. Behav. Pediatr.* **11,** 105–111.
4. Parker, K. J., Buckmaster, C. L., Schatzberg, A. F. & Lyons, D. M. (2004) *Arch. Gen. Psychiatry* **61,** 933–941.
5. Meaney, M. J. (2001) *Annu. Rev. Neurosci.* **24,** 1161–1192.
6. Levine, S. (1957) *Science* **126,** 405.
7. Smotherman, W. P. & Bell, R. W. (1980) in *Maternal Influences and Early Behavior*, eds. Bell, R. W. & Smotherman, W. P. (Spectrum, New York), pp. 201–210.
8. Lee, M. H. & Williams, D. I. (1974) *Anim. Behav.* **22,** 679–681.
9. Liu, D., Diorio, J., Tannenbaum, B., Caldji, C., Francis, D., Freedman, A., Sharma, S., Pearson, D., Plotsky, P. M. & Meaney, M. J. (1997) *Science* **277,** 1659–1662.
10. Rosenberg, K. M., Denenberg, V. H. & Zarrow, M. X. (1970) *Anim. Behav.* **18,** 138–143.
11. Rosenfeld, P., Suchecki, D. & Levine, S. (1992) *Neurosci. Biobehav. Rev.* **16,** 553–568.
12. Schmidt, M., Enthoven, L., van der Mark, M., Levine, S., de Kloet, E. R. & Oitzl, M. S. (2003) *Int. J. Dev. Neurosci.* **21,** 125–132.
13. Huot, R. L., Plotsky, P. M., Lenox, R. H. & McNamara, R. K. (2002) *Brain Res.* **950,** 52–63.
14. Bowman, R. E. & Wolf, R. C. (1965) *Proc. Soc. Exp. Biol. Med.* **119,** 133–135.
15. Pryce, C. R., Palme, R. & Feldon, J. (2002) *J. Clin. Endocrinol. Metab.* **87,** 691–699.
16. Buss, K. A., Schumacher, J. R., Dolski, I., Kalin, N. H., Goldsmith, H. H. & Davidson, R. J. (2003) *Behav. Neurosci.* **117,** 11–20.
17. Gunnar, M. R. (1992) *Pediatrics* **90,** 491–497.
18. Coe, C. L., Glass, J. C., Wiener, S. G. & Levine, S. (1983) *Psychoneuroendocrinology* **8,** 401–409.
19. Levine, S. & Mody, T. (2003) *Neurosci. Biobehav. Rev.* **27,** 83–89.
20. Jordan, T. C., Hennessy, M. B., Gonzalez, C. A. & Levine, S. (1985) *Physiol. Behav.* **34,** 489–493.
21. Brent, L., Koban, T. & Ramirez, S. (2002) *Biol. Psychiatry* **52,** 1047–1056.
22. Baron, R. M. & Kenny, D. A. (1986) *J. Pers. Soc. Psychol.* **51,** 1173–1182.
23. Kraemer, H. C., Stice, E., Kazdin, A., Offord, D. & Kupfer, D. (2001) *Am. J. Psychiatry* **158,** 848–856.
24. Lyons, D. M., Kim, S., Schatzberg, A. F. & Levine, S. (1998) *Dev. Psychobiol.* **32,** 285–291.
25. Champoux, M., Hwang, L., Lang, O. & Levine, S. (2001) *Psychoneuroendocrinology* **26,** 461–477.
26. Lyons, D. M., Yang, C., Eliez, S., Reiss, A. L. & Schatzberg, A. F. (2004) *J. Neurosci.* **24,** 3655–3662.
27. Denenberg, V. H. (1999) *Dev. Psychobiol.* **34,** 1–3.
28. Suomi, S. J. (1997) *Br. Med. Bull.* **53,** 170–184.
29. Kaler, S. R. & Freeman, B. J. (1994) *J. Child Psychol. Psychiatry* **35,** 769–781.
30. Hopf, S., Hartmann-Wiesner, E., Kuhlmorgen, B. & Mayer, S. (1974) *Folia Primatologica* **21,** 225–249.
31. Cohen, S. & Wills, T. A. (1985) *Psychol. Bull.* **98,** 310–357.
32. Parker, K. J., Buckmaster, C. L., Justus, K. R., Schatzberg, A. F. & Lyons, D. M. (2005) *Biol. Psychiatry* **57,** 848–855.
33. Young, E. A. & Altemus, M. (2004) *Ann. N.Y. Acad. Sci.* **1021,** 124–133.
34. Kitay, J. I. (1961) *Endocrinology* **68,** 818–824.
35. Handa, R. J., Burgess, L. H., Kerr, J. E. & O'Keefe, J. A. (1994) *Horm. Behav.* **28,** 464–476.
36. Boinski, S. & Fragaszy, D. M. (1989) *Anim. Behav.* **37,** 415–428.
37. Lyons, D. M., Ha, C. M. & Levine, S. (1995) *Horm. Behav.* **29,** 177–190.
38. Vogt, J. L. & Levine, S. (1980) *Physiol. Behav.* **24,** 829–832.
39. Mendoza, S. P., Lowe, E. L., Resko, J. A. & Levine, S. (1978) *Physiol. Behav.* **20,** 515–522.

1998 Nature America Inc. • http://medicine.nature.com

Neurogenesis in the adult human hippocampus

PETER S. ERIKSSON[1,4], EKATERINA PERFILIEVA[1], THOMAS BJÖRK-ERIKSSON[2], ANN-MARIE ALBORN[1], CLAES NORDBORG[3], DANIEL A. PETERSON[4] & FRED H. GAGE[4]

Department of Clinical Neuroscience, Institute of Neurology[1], Department of Oncology[2], Department of Pathology[3], Sahlgrenska University Hospital, 41345 Göteborg, Sweden

[4]Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

Correspondence should be addressed to F.H.G.

The genesis of new cells, including neurons, in the adult human brain has not yet been demonstrated. This study was undertaken to investigate whether neurogenesis occurs in the adult human brain, in regions previously identified as neurogenic in adult rodents and monkeys. Human brain tissue was obtained postmortem from patients who had been treated with the thymidine analog, bromodeoxyuridine (BrdU), that labels DNA during the S phase. Using immunofluorescent labeling for BrdU and for one of the neuronal markers, NeuN, calbindin or neuron specific enolase (NSE), we demonstrate that new neurons, as defined by these markers, are generated from dividing progenitor cells in the dentate gyrus of adult humans. Our results further indicate that the human hippocampus retains its ability to generate neurons throughout life.

1998 Nature America Inc. • http://medicine.nature.com

Loss of neurons is thought to be irreversible in the adult human brain, because dying neurons cannot be replaced. This inability to generate replacement cells is thought to be an important cause of neurological disease and impairment. In most brain regions, the generation of neurons is generally confined to a discrete developmental period. Exceptions are found in the dentate gyrus and the subventricular zone of several species that have been shown to generate new neurons well into the postnatal and adult period[1–6]. Granule neurons are generated throughout life from a population of continuously dividing progenitor cells residing in the subgranular zone of the dentate gyrus in the rodent brain[5]. 'Newborn' neurons generated from these progenitor cells migrate into the granule cell layer, differentiate, extend axons and express neuronal marker proteins[7–10].

We examined whether progenitor cells reside in the adult human hippocampus and whether new neurons are born within the dentate gyrus of the adult human brain. Postmortem tissue from the hippocampus and the subventricular zone of caudate nucleus was obtained from cancer patients (n = 5) who received one intravenous infusion (250 mg; 2.5 mg/ml, 100 ml) of bromodeoxyuridine (BrdU) for diagnostic purposes[11]. One patient diagnosed with a similar type and location of cancer, but without BrdU treatment, was included as a control. A thymidine analog, BrdU is incorporated into the DNA of dividing cells and can be detected immunohistochemically in their progeny[5,12,13].

Cell genesis and survival in the adult human dentate gyrus

The number of surviving labeled, proliferating progenitors was

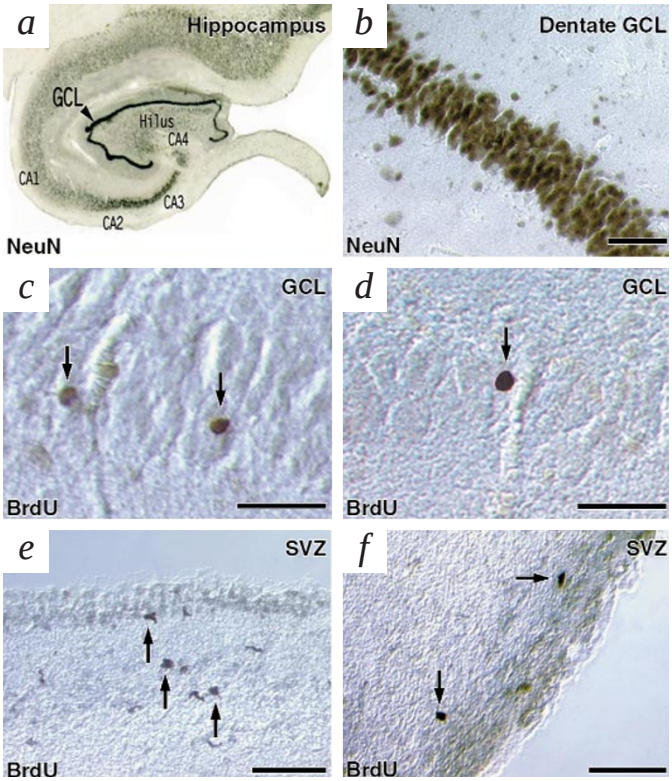


Fig. 1 Newly generated cells can be detected in the adult human brain in patients previously treated with BrdU. ***a***, The hippocampal region of the adult human brain immunoperoxidase-stained for the neuronal marker NeuN. ***b***, The hippocampal dentate gyrus granule cell layer (GCL) visualized with immunoperoxidase staining for NeuN. ***c***, Differential interference contrast photomicrograph showing BrdU-labeled nuclei (arrows) in the dentate granule cell layer (GCL). ***d***, Differential interference contrast photomicrograph showing a BrdU-labeled nucleus (arrow) in the human dentate GCL. BrdU-positive nuclei have a rounded appearance and resemble the chromatin structure of mature granule cells and are found within the granule cell layer. ***e***, Differential interference contrast photomicrograph showing BrdU-positive cells (arrows) adjacent to the ependymal lining in the subventricular zone of the human caudate nucleus. Cells with elongated nuclei resembling migrating cells are in the rat subventricular zone (SVZ). ***f***, Differential interference contrast photomicrograph showing BrdU-positive cells (arrows) with round to elongated nuclei in the subventricular zone of the human caudate nucleus. All scale bars represent 50 μm.

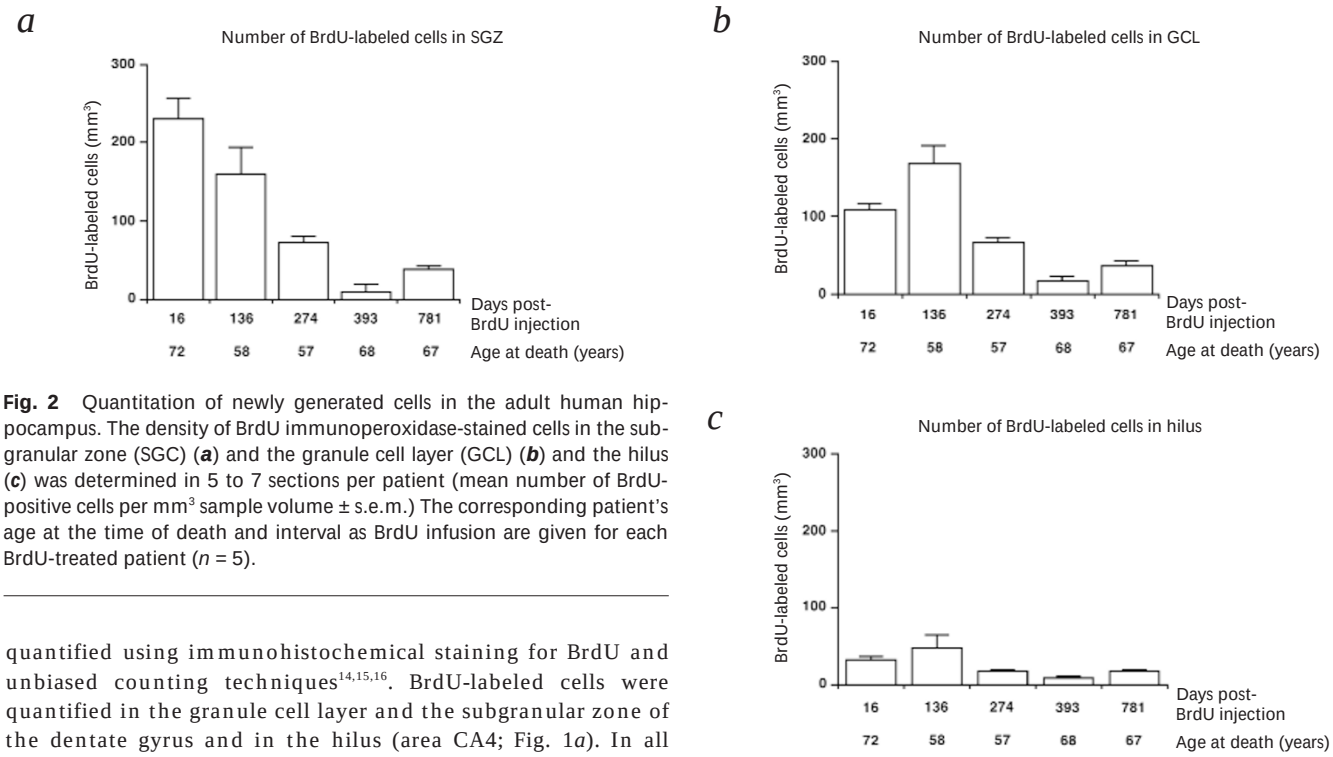


Fig. 2 Quantitation of newly generated cells in the adult human hippocampus. The density of BrdU immunoperoxidase-stained cells in the subgranular zone (SGC) (***a***) and the granule cell layer (GCL) (***b***) and the hilus (***c***) was determined in 5 to 7 sections per patient (mean number of BrdU-positive cells per mm^3 sample volume ± s.e.m.) The corresponding patient's age at the time of death and interval as BrdU infusion are given for each BrdU-treated patient (n = 5).

quantified using immunohistochemical staining for BrdU and unbiased counting techniques[14,15,16]. BrdU-labeled cells were quantified in the granule cell layer and the subgranular zone of the dentate gyrus and in the hilus (area CA4; Fig. 1*a*). In all BrdU-treated patients, the granule cell layer contained BrdU-positive, round- to oval-shaped nuclei with the typical morphology of granule cell neuron nuclei (Fig. 1*b–d*). The morphological appearance and localization of these nuclei were similar to those of the nuclei of cells found in the dentate gyrus of adult mouse, rat and marmoset monkey[5,6,17]. No BrdU-immunoreactive profiles could be detected in the control patient that had not received BrdU. The presence of BrdU-positive nuclei in a brain region that is neurogenic in many other mammalian and primate species indicates that proliferating neural progenitor cells are present in the adult human dentate gyrus.

The number of BrdU-positive cells in the granule cell layer, the subgranular zone and hilus varied between individuals, as may be expected because of the varying post-infusion intervals and different ages of the patients (Fig. 2*a–c*). There is an apparent decline in the number of cells that are detected in the patients that had the longest interval between BrdU treatment and histological assessment. This decline may indicate a progressive death of the newly generated cells over time. However, the small sample weakens any quantitative conclusion.

BrdU-labeled cells co-express neuronal markers

To determine the cell fate of the BrdU-immunoreactive cells, we did triple immunofluorescent labeling for BrdU and cell-specific markers, including glial fibrillary acidic protein (GFAP), a marker for astroglia, and one of the neuronal markers, NeuN[18,19], calbindin[20] or neuron specific enolase[21] (NSE). Confocal microscopy was used to determine the phenotype of the BrdU-positive cells (Figs. 3 and 4). No evidence for cross-reactivity between GFAP-immunoreactivity and any of the three neuronal markers was detected in these sections. GFAP-immunopositive processes were observed to surround neurons and their processes, but did not co-express in the same cell (Fig. 3*e–h*). Autofluorescence was observed in this tissue, but the nonspecific emission artifact was easily detected by its contribution to multiple wavelengths (Fig. 3*a–d*). We took care to evaluate potential sources of autofluorescence in the human tissue by directly imaging untreated sections and sections from the control patient both with and without the pretreatment for BrdU-immunohistochemical detection in the absence of antibodies against BrdU. The autofluorescence was present even in completely untreated sections and was restricted to the cytoplasm of neuronally marked cells or to endothelial cells and red blood cells and could be clearly distinguished from BrdU-immunoreactive profiles (Fig. 3). The majority of the NeuN-positive neurons that double-labeled with BrdU were located within or near the granule cell layer and had the morphological characteristics of granule cell neurons: round or oval nuclei with small- to medium-sized cell bodies (Fig. 3). Confocal microscopic Z-series analyses of BrdU-positive nuclei unambiguously identified double-labeling with NeuN and demonstrated the existence of newly generated (BrdU-labeled) granule cell neurons in the dentate gyrus of the adult human brain (Fig. 3*d–h*). The phenotypic expression of these BrdU-positive neurons in adult humans was equivalent to the expression found in adult rodents[5,10,12,17] (Fig. 3*i–l*). The average fraction of BrdU-NeuN double-labeled cells in the granule cell layer was 22.0 ± 2.4% among the BrdU-treated patients.

Neurons double-labeled with BrdU and NSE could also be detected in all BrdU-treated subjects (Fig. 4*e–f*). The average fraction of BrdU-labeled cells that were also NSE-labeled in the granule cell layer was 22.7 ± 2.8%. Cells that double-labeled for BrdU and calbindin were also observed and provided further support for the occurrence of neurogenesis in the adult SGZ, GCL and hilus (Fig. 4*a–d*). The average fraction of BrdU-labeled cells that were also calbindin-positive in the granule cell layer was 7.9 ± 2.2% among the BrdU-treated patients. The dentate gyrus from all individuals contained a fraction of BrdU-positive cells that were also GFAP-positive, 18.1 ± 1.8%, and had the typical morphology of star-shaped glial cells with small, irregularly shaped nuclei and small cell bodies with thin, GFAP-positive processes (Fig. 4*c–d*). The dentate gyrus from all subjects contained BrdU-

1998 Nature America Inc. • http://medicine.nature.com

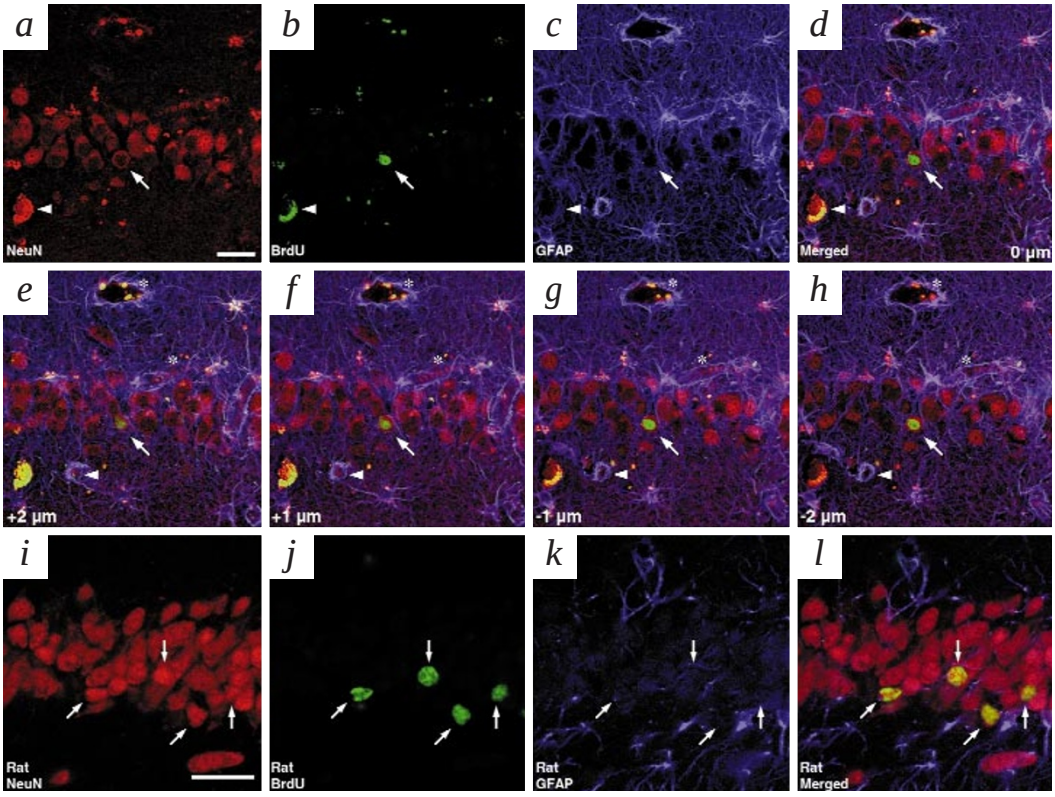


Fig. 3 Newly generated cells in the adult human dentate gyrus can express a neuronal phenotype. Simultaneous detection of immunofluorescent labels for NeuN (***a***; scale bar represents 25 µm), BrdU (***b***) and GFAP (***c***) for detection of astrocytes and a merge of these signals in x-, y- and z-registration (***d***) examined by confocal microscopy showed that BrdU-labeled nuclei could specifically co-express NeuN without expressing GFAP (arrows in *a*–*d*). In addition to specific BrdU-labeling, some neurons contained nonspecific fluorescence in the green and red emission, reflecting their accumulation of lipofuchsin granules (arrowheads in *a*–*d*). Red blood cells and endothelial cells, present in several small blood vessels, also emit nonspecific green and red fluorescence (small blood vessel indicated by arrowheads, larger blood vessels indicated by asterisks in *e*–*h*). The specificity of BrdU-NeuN co-expression in three-dimensions is demonstrated by a series of focal planes above (***e***,***f***) and below (***g***, ***h***) the focal plane shown in *d* (arrows indicate the same cell as in *d*). The appearance of adult BrdU-positive neurons in the human dentate gyrus is equivalent to BrdU-labeled neurons in the adult rat dentate gyrus (***i***–***l***; scale bar represents 25 µm). The rat tissue (*i*–*l*) is at a different magnification than the human tissue (*a*–*h*).

positive cells that were immunonegative for both neuronal and glial markers. These immunonegative cells likely represent quiescent undifferentiated cells, newborn cells of a phenotype not examined here and/or a pool of asymmetrically dividing progenitor cells. These cells were characterized by small, round-to-oval BrdU-positive nuclei and the absence of cell-specific immunoreactivity.

To further confirm the presence of neurogenesis, we double-labeled using antibodies against BrdU and either NSE, calbindin or NeuN with chromagens for brightfield optics (alkaline phosphatase (Vector Blue) for BrdU and 3-amino-9-ethylcarbazole (AEC) for the neuronal markers). Although confocal microscopy could not be used for imaging, the brightfield chromagens had the advantage of not fading with examination and not contributing autofluorescence to the image. Examination of brightfield staining confirmed that BrdU-positive cells in the adult human hippocampus could express a neuronal phenotype (Fig. 5*a*–*e*) morphologically indistinguishable from adult rodent BrdU-positive neurons (Fig. 5*f* and *g*), strongly supporting our conclusion that neurogenesis occurs in the dentate granule cell layer of the adult human brain.

BrdU labeling in the subventricular zone of adult human brain

Another neurogenic region, the subventricular zone (SVZ) adjacent to the caudate nucleus, was examined for the presence of BrdU-positive cells. Tissue samples from all BrdU-treated patients contained BrdU-positive cells within the SVZ (Fig. 1*e*–*f*). BrdU-positive cells did not co-express the cell-specific markers GFAP and NeuN (data not shown). The morphology of BrdU-labeled nuclei within the SVZ was small and round-to-oval, resembling that of progenitor cells in the rat SVZ (ref. 5). This finding supports the idea that the human SVZ contains progenitor cells and that these cells are required to migrate from the SVZ before they differentiate[22]. We were unable to detect any BrdU-immunoreactive cells in tissue from the control patient who had not received BrdU treatment, supporting the interpretation that the BrdU staining we report here reflects the persistence in the adult human brain of cell genesis.

Discussion

Our study demonstrates that cell genesis occurs in human brains and that the human brain retains the potential for self-renewal throughout life. Although earlier studies in adult primates have been unsuccessful in showing neurogenesis in the dentate gyrus[23,24], a recent report has demonstrated neurogenesis in three-year-old marmoset monkeys[6]. Although the number of BrdU-labeled cells entering the neuronal lineage seems to be lower in the human hippocampus than in marmosets, those monkeys[6] were considerably younger, even in relative terms, than the humans examined here (average age of 64.4 ± 2.9 years). Therefore, we conclude that, as in rodents[5,25], neurogenesis in the human dentate gyrus continues throughout life.

Although our results demonstrate that cells in the adult brain undergo cell division and that some of the newly generated cells can survive and differentiate into cells with morphological and phenotypic characteristics of neurons, we have not proven that these newly generated cells are functional. We also do not yet know the biological significance of cell genesis in the adult human brain. However, this does provide a basis to investigate a newly discovered type of 'neuroplasticity' in humans, one based

1998 Nature America Inc. • http://medicine.nature.com

1998 Nature America Inc. • http://medicine.nature.com

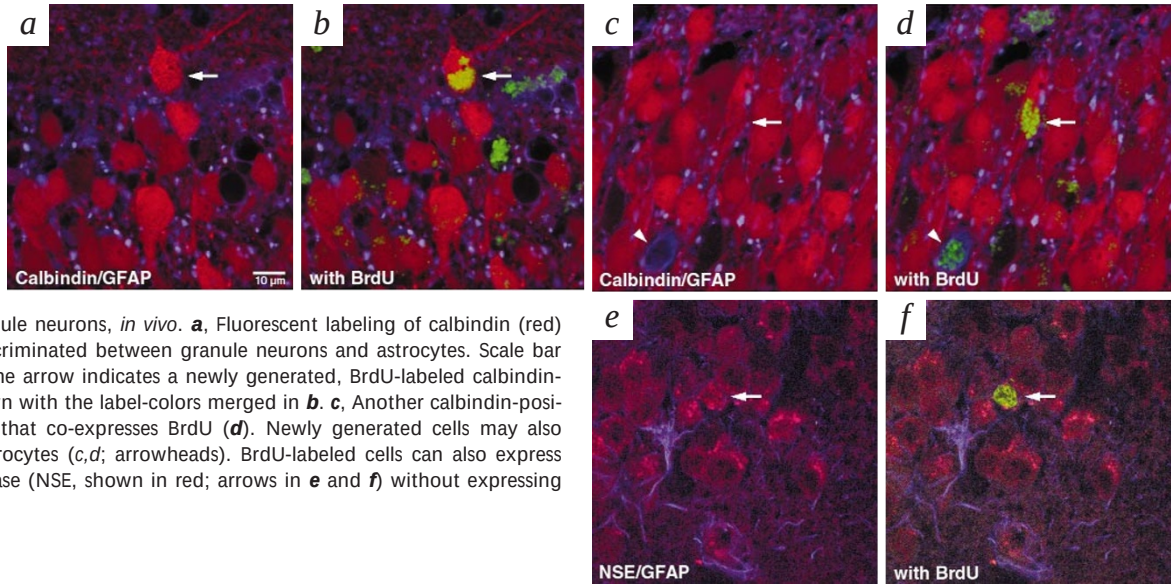


Fig. 4 Newly generated cells in the adult human dentate gyrus can express additional neuronal phenotypes. The calcium-binding protein, calbindin, is expressed by certain neuronal populations, including dentate granule neurons, *in vivo*. ***a***, Fluorescent labeling of calbindin (red) and GFAP (blue) discriminated between granule neurons and astrocytes. Scale bar represents 10 μm. The arrow indicates a newly generated, BrdU-labeled calbindin-positive neuron shown with the label-colors merged in ***b***. ***c***, Another calbindin-positive neuron (arrow) that co-expresses BrdU (***d***). Newly generated cells may also differentiate into astrocytes (*c,d*; arrowheads). BrdU-labeled cells can also express neuron specific enolase (NSE, shown in red; arrows in ***e*** and ***f***) without expressing GFAP (blue).

on addition of neurons, that has not been previously considered. Studies in rodents have shown that the adult hippocampus contains progenitor cells that can be expanded *in vitro* and grafted back into the adult brain, where they can respond to regional cues by differentiation into site-specific phenotypes, including neurons[26,27]. The presence of progenitor cells in the human dentate gyrus, reported here, indicates that these cells also may be used for *in vitro* and *in vivo* studies of cell differentiation and possibly subsequent transplantation studies. Furthermore, environmental stimulation can influence the rate of neurogenesis in the adult and senescent rodent dentate gyrus[12,17]. The potential to regulate human neurogenesis should prove to be an interesting area of investigation.

Methods

Autopsy material. Human hippocampal tissue was obtained at autopsy with the full consent of each family. All patients were diagnosed with squamous cell carcinomas at the base of the tongue, in the larynx or in the pharynx. All patients received bromodeoxyuridine (BrdU) (250 mg) dissolved in saline and given as an intravenous infusion (2.5 mg/ml, 100 ml). The BrdU was given to the patients to assess the proliferative activity of the tumor cells, expressed as BrdU-labeling index. No signs of macroscopic or microscopic metastases were found in autopsy material from the cerebrum in any of the patients. No anti-cancer therapy was administered before or during BrdU administration to any of the patients.

Tissue preparation. The hippocampal formation and ventricular zone were dissected out and post-fixed in 4% paraformaldehyde for 24 hours and then transferred into 30% sucrose solution until equilibrated. The hippocampi were sectioned (slices 40 μm in thickness) in the coronal plane on a sliding microtome and stored at –20 °C in a cryoprotecting buffer containing 25% ethylene glycol, 25% glycerin and 0.05 M phosphate buffer. For comparison, sections derived from BrdU-injected adult rats and mice that had been intracardially perfused with 4% paraformaldehyde were processed in parallel with the human tissue.

Histology. Immunohistochemical detection of BrdU requires a pre-treatment of the sections to denature DNA (ref. 5). All staining was done on free-floating sections and the blocking buffer contained both 3% normal donkey serum and 3% normal human serum (Sigma). For sections stained only for BrdU, a

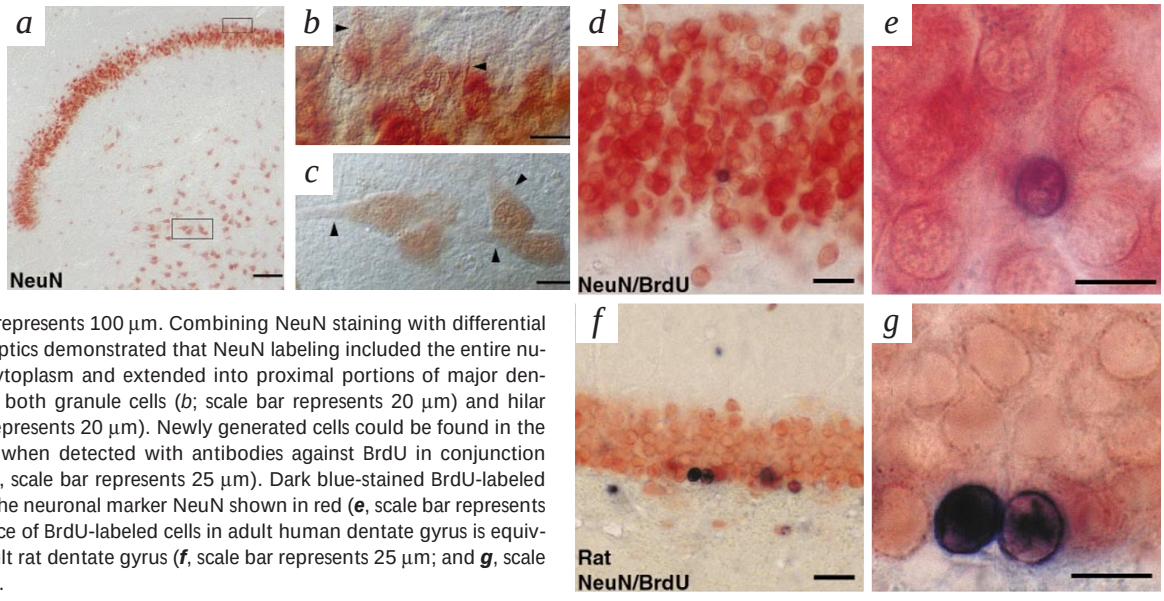


Fig. 5 Brightfield double-immunohistochemical demonstration of neuronal phenotype. ***a***, Staining with the neuronal marker NeuN labels both dentate granule cells (top inset, shown in ***b***) and hilar neurons (bottom inset, shown in ***c***). Scale bar represents 100 μm. Combining NeuN staining with differential interference contrast optics demonstrated that NeuN labeling included the entire nucleus and perikaryal cytoplasm and extended into proximal portions of major dendrites (arrowheads) in both granule cells (*b*; scale bar represents 20 μm) and hilar neurons (*c*, scale bar represents 20 μm). Newly generated cells could be found in the dentate granule layer when detected with antibodies against BrdU in conjunction with NeuN staining (***d***, scale bar represents 25 μm). Dark blue-stained BrdU-labeled nuclei can co-express the neuronal marker NeuN shown in red (***e***, scale bar represents 10 μm). The appearance of BrdU-labeled cells in adult human dentate gyrus is equivalent to that of the adult rat dentate gyrus (***f***, scale bar represents 25 μm; and ***g***, scale bar represents 10 μm.).

1998 Nature America Inc. • http://medicine.nature.com

mouse-anti-BrdU antibody (Boehringer; diluted 1:400) was used in combination with the avidin-biotin complex method using a biotinylated donkey anti-mouse-IgG antibody (Vector Laboratories, Burlingame, California; diluted 1:167) and reacted with a diaminobenzidine (DAB) chromagen.

Immunofluorescent double- and triple-labeling was done as described[5,12]. For multiple immunostaining, BrdU was detected with a rat anti-BrdU antibody (Harlan Sera-Lab, Loughborough, England; diluted 1:500) and visualized with a FITC-conjugated secondary donkey anti-rat antibody (Jackson ImmunoResearch, West Grove, Pennsylvania; diluted 1:250). For neuronal phenotype markers, sections were incubated with one of the following antisera: rabbit anti-calbindin antiserum (SWant, Bellinzona, Switzerland; diluted 1:1,000), or mouse anti-NeuN antiserum (from R. Mullen[18]; diluted 1:20), or rabbit anti-NSE antiserum (Polysciences, Warrington, Pennsylvania; diluted 1:800). These neuronal markers were visualized by using the species-appropriate Cy3-conjugated secondary antibody (Jackson ImmunoResearch, West Grove, Pennsylvania; diluted 1:250). GFAP was detected in the same sections using a guinea pig anti-GFAP antiserum (Advanced Immunochemicals, Long Beach, California; diluted 1:250) and visualized using a Cy5-conjugated donkey anti-guinea pig antibody (Jackson ImmunoResearch, West Grove, Pennsylvania; diluted1:250).

For double immunostaining using brightfield chromagens, sections were incubated with a pooled solution of rat anti-BrdU (Harlan Sera-Lab, Loughborough, England; diluted 1:500) and mouse anti-NeuN (from R. Mullen[18]; diluted 1:100) antisera. After being rinsed and blocked, sections were incubated first with a biotinylated donkey anti-rat antibody (Jackson ImmunoResearch, West Grove, Pennsylvania; diluted 1:250), followed by incubation with an alkaline phosphatase avidin–biotin substrate and then reaction with the blue chromagen (Vector Blue; Vector Laboratories, Burlingame, California). After being rinsed and blocked further, the sections then were incubated with a biotinylated donkey anti-mouse antibody (Jackson ImmunoResearch, West Grove, Pennsylvania; diluted 1:250) followed by incubation with a peroxidase avidin-biotin substrate and then reaction with the red chromagen (3-amino-9-ethylcarbazole or AEC; Vector Laboratories, Burlingame, California). Sections were mounted and coverslipped with an aqueous mounting medium.

Fluorescent signals were imaged using a confocal laser scanning microscope equipped with a krypton/argon mixed gas laser (Bio-Rad MRC1024; Bio-Rad, Richmond, California) using individual collection of wavelengths to minimize artifactual detection of non-specific fluorescent emission. At least 20 BrdU-positive cells within multiple immunofluorescently stained sections from each patient were examined on the confocal microscope to determine the phenotype of BrdU-labeled cells. Brightfield images were derived from conventional photomicrographs digitized by using a slide scanner. Figures were composed by using Adobe Photoshop (Adobe Systems, Mountain View, California) with image adjustments confined to creating an equivalent signal distribution between panels.

Quantitation. The numbers of BrdU-positive cells in the granule cell layer, the subgranular layer and the hilus, and their corresponding sample volumes were determined in five to seven immunoperoxidase-stained coronal sections 40 μm in thickness, at least 240 μm apart, from each patient. Area estimations were done on an IBAS image analysis system. The section thickness of 40 μm (microtome setting) was used in the dissector estimation of volume. The number of BrdU-positive cells was counted within the granule cell layer (GCL) and two cell diameters below the granule cell layer, ignoring the cells in the uppermost focal plane and focusing through the thickness of the section (optical disector principle; refs. 14,15,16) to avoid oversampling errors. The subgranular zone (SGZ) was defined as the area from one cell diameter within the GCL from the hilus-GCL border and two cell diameters below the hilus-GCL border (Fig. 1). Quantitative analysis of BrdU immunoperoxidase-stained sections was done on a Nikon microscope equipped with a video camera. The results are expressed as BrdU-positive cells per sample volume per section.

Acknowledgments

We thank G. Kempermann, T. Palmer, E. Brandon, M.-C. Senut, K. Sakurada, L. Chehabeddine and M. L. Gage for their comments, and H. van Praag for the contribution of rat tissue. In addition, we thank L. Kitabayashi for technical assistance. This study was supported by grants from the Swedish Medical Research Council (project no. K98-12X-12535-01A), Faculty of Medicine, University of Göteborg, the Gunvor and Josef Anérs Stiftelse, the John and Brit Wennerströms Stiftelse for Neurologisk Forskning, the Rune and Ulla Amlövs Stiftelse for Neurologisk and Reumatologisk Forskning, NHR-fonden, Stiftelsen Göteborgs MS förenings forsknings och byggnadsfond, Stiftelsen Handlanden Hjalmar Svenssons Forskningsfond, Göteborgs Läkaresällskap, Hjärnfonden, The Swedish Society of Medicine, Stiftelsen Lars Hiertas Minne, Stiftelsen Assar Gabrielssons Fond and the Edit Jacobssons Fond and from NIA and NINDS and the Alzheimer's Association (F.H.G.) and the American Federation for Aging Research (D.A.P.).

RECEIVED 9 SEPTEMBER; ACCEPTED 13 OCTOBER 1998

1. Altman, J. &. Das, G.D. Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. *J. Comp. Neurol.* **124**, 319–335 (1965).
2. Altman, J. & Das, G.D. Postnatal neurogenesis in the guinea-pig. *Nature* **214**, 1098–1101 (1967).
3. Caviness, V.S. Time of neuron origin in the hippocampus and dentate gyrus of normal and reeler mutant mice: an autoradiographic analysis. *J. Comp. Neurol.* **151**, 113–120 (1973).
4. Gueneau, G., Privat, A., Drouet, J. & Court, L. Subgranular zone of the dentate gyrus of young rabbits as a secondary matrix. A high-resolution autoradiographic study. *Dev. Neurosci.* **5**, 345–358 (1982).
5. Kuhn, H.G., Dickinson-Anson, H. & Gage, F.H. Neurogenesis in the dentate gyrus of the adult rat: Age-related decrease of neuronal progenitor proliferation. *J. Neurosci.* **16**, 2027–2033 (1996).
6. Gould, E., Tanapat, P., McEwen, B.S., Flugge, G. & Fuchs, E. Proliferation of granule cell precursors in the dentate gyrus of adult monkeys is diminished by stress. *Proc. Natl. Acad. Sci. USA* **95**, 3168–3171 (1998).
7. Kaplan, M.S. & Bell, D.H. Mitotic neuroblasts in the 9-day-old and 11-month-old rodent hippocampus. *J. Neurosci.* **4**, 1429–1441 (1984).
8. Kaplan, M.S. & Hinds, J.W. Neurogenesis in the adult rat: electron microscopic analysis of light radioautographs. *Science* **197**, 1092–1094 (1977).
9. Stanfield, B.B. & Trice, J.E. Evidence that granule cells generated in the dentate gyrus of adult rats extend axonal projections. *Exp. Brain Res.* **72**, 399–406 (1988).
10. Cameron, H.A., Woolley, C.S., McEwen, B.S. & Gould, E. Differentiation of newly born neurons and glia in the dentate gyrus of the adult rat. *Neuroscience* **56**, 337–344 (1993).
11. Dolbeare, F. Bromodeoxyuridine: a diagnostic tool in biology and medicine, Part I: Historical perspectives, histochemical methods and cell kinetics. *Histochem. J.* **27**, 339–369 (1995).
12. Kempermann, G., Kuhn, H.G. & Gage, F.H. Genetic influence on neurogenesis in the dentate gyrus of adult mice. *Proc. Natl. Acad. Sci. USA* **94**, 10409–10414 (1997).
13. del Rio, J.A. & Soriano, E. Immunocytochemical detection of 5'-bromodeoxyuridine incorporation in the central nervous system of the mouse. *Dev. Brain Res.* **49**, 311–317 (1989).
14. Gundersen, H.J. *et al*. The new stereological tools: disector, fractionator, nucleator and point sampled intercepts and their use in pathological research and diagnosis. *APMIS* **96**, 857–881 (1988).
15. West, M.J. Regionally specific loss of neurons in the aging human hippocampus. *Neurobiol. Aging* **14**, 287–293 (1993).
16. Coggeshall, R.E. & Lekan, H.A. Methods for determining numbers of cells and synapses: a case for more uniform standards of review. *J. Comp. Neurol.* **364**, 6–15 (1996).
17. Kempermann, G., Kuhn, H.G. & Gage, F.H. More hippocampal neurons in adult mice living in an enriched environmen *Nature* **386**, 493–495 (1997).
18. Mullen, R.J., Buck, C.R. & Smith, A.M. NeuN, a neuronal specific nuclear protein in vertebrates. *Development* **116**, 201–211 (1992).
19. Wolf, H.K. *et al*. NeuN: A useful neuronal marker for diagnostic histopathology. *J. Histochem. Cytochem.* **44**, 1167–1171 (1996).
20 Sloviter, R.S. Calcium-binding protein (calbindin-D28k) and parvalbumin immunocytochemistry: localization in the rat hippocampus with specific reference to the selective vulnerability of hippocampal neurons to seizure activity. *J. Comp. Neurol.* **280**, 183–196 (1989).
21. Schmechel, D., Marango, P.J., Zis, A.P., Brightman, M. & Goodwin, F.K. Brain enolases as specific markers of neuronal and glial cells. *Science* **199**, 313–315 (1978).
22. Kirschenbaum, B. *et al*. *In vitro* neuronal production and differentiation by precursor cells derived from the adult human forebrain. *Cereb. Cortex* **4**, 576–589 (1994).
23. Rakic, P. Limits of neurogenesis in primates. *Science* **227**, 1054–1056 (1985).
24. Eckenhoff M.F. & Rakic, P. Nature and fate of proliferative cells in the hippocampal dentate gyrus during the life span of the rhesus monkey. *J. Neurosci.* **8**, 2729–2747 (1988).
25. Kempermann, G., Kuhn, H.G. & Gage, F.H. Experience-induced neurogenesis in the senescent dentate gyrus. *J. Neurosci.* **18**, 3206–3212 (1998).
26. Gage, F.H. *et al*. Survival and differentiation of adult neuronal progenitor cells transplanted to the adult brain. *Proc. Natl. Acad. Sci. USA* **92**, 11879–11883 (1995).
27. Suhonen, J.O., Peterson, D.A., Ray, J. & Gage, F.H. Differentiation of adult hippocampus-derived progenitors into olfactory neurons in vivo. *Nature* **383**, 624–627 (1996).

1998 Nature America Inc. • http://medicine.nature.com

Neuron
Article

NMDA Receptors in Dopaminergic Neurons Are Crucial for Habit Learning

Lei Phillip Wang,[1] Fei Li,[1,2] Dong Wang,[1] Kun Xie,[1] Deheng Wang,[1] Xiaoming Shen,[2] and Joe Z. Tsien[1,*]
[1]Brain and Behavior Discovery Institute and Department of Neurology, Georgia Health Sciences University, Augusta, GA 30912, USA
[2]Department of Developmental and Behavioral Pediatrics of Shanghai Children's Medical Center & Shanghai Key Laboratory of Children's Environmental Health, XinHua Hospital affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China
*Correspondence: jtsien@georgiahealth.edu
DOI 10.1016/j.neuron.2011.10.019

SUMMARY

Dopamine is crucial for habit learning. Activities of midbrain dopaminergic neurons are regulated by the cortical and subcortical signals among which glutamatergic afferents provide excitatory inputs. Cognitive implications of glutamatergic afferents in regulating and engaging dopamine signals during habit learning, however, remain unclear. Here, we show that mice with dopaminergic neuron-specific NMDAR1 deletion are impaired in a variety of habit-learning tasks, while normal in some other dopamine-modulated functions such as locomotor activities, goal-directed learning, and spatial reference memories. In vivo neural recording revealed that dopaminergic neurons in these mutant mice could still develop the cue-reward association responses; however, their conditioned response robustness was drastically blunted. Our results suggest that integration of glutamatergic inputs to DA neurons by NMDA receptors, likely by regulating associative activity patterns, is a crucial part of the cellular mechanism underpinning habit learning.

INTRODUCTION

Many acts, after repetitive practice, would transform from being goal-directed to automated habits, which can be carried out efficiently and subconsciously. Habits help to free up the cognitive loads on routine procedures and allow us to focus on new situations and tasks. Despite breakthroughs unveiling participation of different anatomical structures in habit formation (Knowlton et al., 1996; Yin and Knowlton, 2006), the underpinning physiological mechanisms and how different network circuitries integrate relevant information remain unclear.

Dopamine (DA) is an important regulator of synaptic plasticity, especially in the basal ganglia, a structure essential for habit learning. In both human patients (Fama et al., 2000; Knowlton et al., 1996) and rodents (Faure et al., 2005), habit learning is often found impaired following dopaminergic neuron degeneration. Dopamine has thus been postulated as a main modulator in the mechanisms subserving habit learning (Ashby et al., 2010). Despite this importance, the mechanisms modulating dopamine during habit learning have yet to be fully investigated. Studies have shown that habit-learning deficits caused by dopamine deafferentation could not be rescued by simple intrastriatal injections of DA agonists (Faure et al., 2010). These observations suggested that dopamine, the modulator itself, might need to be regulated during normal habit learning. Anatomically, along with cholinergic inputs, glutamatergic afferents from brain structures such as pedunculopontine tegmental nucleus (PPTg), subthalamic nucleus (STN), and prefrontal cortex (PFC) provide main forms of excitatory inputs to the midbrain DA neurons (Grace et al., 2007). NMDA receptors (NMDARs), members of the ionotropic glutamate receptor family, are important regulators of DA neuron activity. First, synaptic plasticity in the glutamatergic afferents to DA neurons depends on NMDARs (Bonci and Malenka, 1999; Overton et al., 1999; Ungless et al., 2001). This plasticity can be modulated by experiences, environmental factors, and psychostimulant drugs (Bonci and Malenka, 1999; Kauer and Malenka, 2007; Saal et al., 2003). Second, iontophoretic administration of NMDAR antagonists, but not AMPAR-selective antagonists, attenuated phasic firing of DA neurons, an activity linked to reward/incentive salience (Schultz, 1998), without changing the frequency of tonic firing (Overton and Clark, 1992). Third, in drug addiction studies, NMDARs in DA neurons are essential for developing nicotine-conditioned place preference (Wang et al., 2010) and likely also involved in cocaine-conditioned place preference (Engblom et al., 2008; Zweifel et al., 2008). Thus, we postulated that modulation of DA neurons by NMDARs might be important in engaging DA neurons in the habit learning. Here, we set out to examine the roles of NMDARs in DA neurons, by generating DA neuron-specific NR1 knockout mice and testing them in a variety of habit-learning paradigms (Devan and White, 1999; Dickinson et al., 1983; Packard et al., 1989; Packard and McGaugh, 1996). In order to understand the cellular mechanisms, we also recorded the DA neurons in these mice using multielectrode in vivo neural-recording techniques (Wang and Tsien, 2011).

RESULTS

Production and Basic Characterization of DA Neuron-Selective NR1 Knockout Mice

These mice, named "DA-NR1-KO," were produced by crossing floxed NR1 (fNR1) mice (Tsien et al., 1996) with *Slc6a3+/Cre*

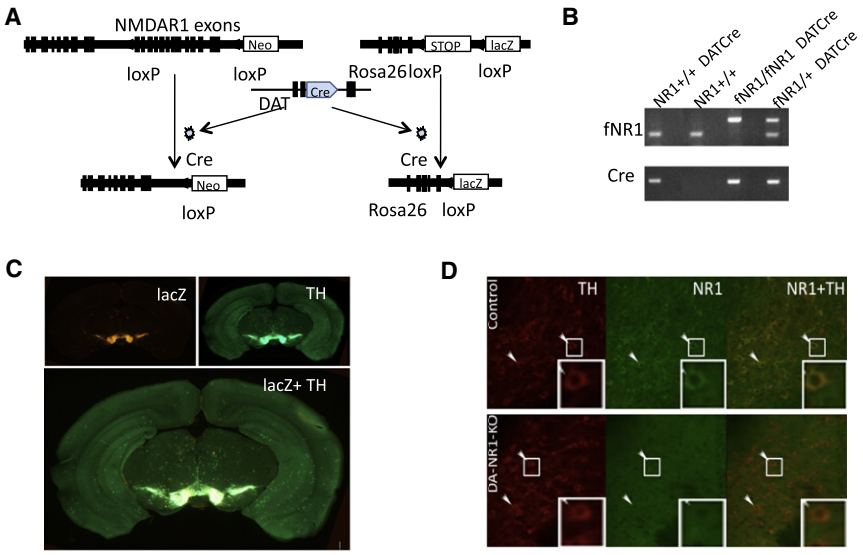


Figure 1. Generation and Characterization of DA-NR1-KO Mice

(A) Breeding scheme for making DA-NR1-KO mice and for DAT-Cre/rosa-stop-lacZ mice.

(B) PCR genotyping of DA-NR1-KO (Cre, fNR1/fNR1) mice and their littermates.

(C) Immunofluorescent staining of DAT-Cre/Rosa-Stop-lacZ mouse brain with lacZ antibody, with anti-TH antibody and overlay of the two stainings.

(D) Immunofluorescent staining of DA-NR1-KO and wild-type control brain with anti-NR1 antibody, anti-TH antibodies and overlay of both stainings. Arrowheads indicate position of some TH-positive neurons. TH signals colocalize with NR1 in the control brain, but not in the DA-NR1-KO brains. Pictures in the larger boxes are the 3-fold enlargement of the pictures shown in the respective small boxes.

transgenic mice that express Cre recombinase under DA transporter promoter (Zhuang et al., 2005) (Figures 1A and 1B). The DA neuron-specific deletion of the NR1 gene was confirmed by both the reporter gene method (Figure 1C) and immunohistochemistry (Figure 1D), which showed that the gene deletion was restricted to the dopaminergic neurons in regions such as the VTA and the substantia nigra. No obvious changes were observed in the expression pattern of tyrosine hydroxylase (TH), the catecholamine neuronal marker, suggesting that there was no obvious loss of dopaminergic neurons (see Figure S1 available online).

DA-NR1-KO mice were born in the expected Mendelian ratios and visually indistinguishable from the controls. Additionally, they were normal in locomotor activities in a novel open field (Figure 2A), in learning the rotarod tests (Figure 2B), in an anxiety test using the elevated plus maze (Figure 2C), and in the novel object recognition tests (Figure 2D). These results showed that many of the behavioral functions that were sensitive to dopamine dysfunctions were preserved in the DA-NR1-KO mice.

DA Neurons in the DA-NR1-KO Showed Normal Tonic Firing, Reduced Phasic Firing, and Reduced Responses to the Reward-Predicting Cue

In order to investigate the impact of NR1 deletion on the cellular properties of DA neurons, we recorded the activities of these neurons in both the DA-NR1-KO mice and wild-type control littermates. Movable bundles of 8 tetrodes (32 channels) were implanted into the ventral midbrain, primarily the VTA. The putative DA neurons were identified based on their firing patterns and their sensitivity to dopamine receptor agonist apomorphine (1 mg/kg, i.p.) at the end of each recording session (Figures 3A–3D).

A total of 14 putative DA neurons from 4 mutant mice and 16 from 6 wild-type controls were recorded and analyzed. Phasic-firing activities or bursting was defined as a spike train beginning with an interspike interval (ISI) smaller than 80 ms and terminating with an ISI greater than 160 ms. Compared with the control neurons, phasic-firing activities were greatly reduced in the DA-NR1-KO neurons. The observed median frequency of phasic firing decreased from 0.78 ± 0.09 Hz in the control DA neurons to 0.36 ± 0.09 Hz in KO DA neurons. (Mann-Whitney U test, $p < 0.01$) (Figure 3E). A significant reduction was also observed in the percentages of spikes fired in phasic activities (34.7% in the controls versus 21.2% in the DA-NR1-KO; Mann-Whitney U test, $p < 0.01$) (Figure 3F). The total firing rate was also reduced in the mutant DA neurons. This appeared to be correlated with reduced burst set rate (5.18 ± 0.59 Hz, control, versus 3.85 ± 0.38 Hz, KO; $r = 0.7719$, Mann-Whitney U test, $p < 0.01$) (Figure 3G). No significant difference was observed in the tonic firing between the mutant and control groups. (4.42 ± 0.44 Hz in control, versus 3.29 ± 0.36 Hz in KO; Mann-Whitney U test, $p > 0.05$) (Figure 3H).

To further evaluate the response of DA neurons in a learning task, mice were trained 40 trials per day in a Pavlovian-conditioning paradigm in which a 5 KHz tone that lasted 1 s proceeded immediately before the delivery of a food pellet. DA neurons from both genotypes were able to associate the tone with phasic firing, but the conditioned responses were much weaker in the DA-NR1-KO group (Figure 4A). Although DA-NR1-KO neurons showed increased firing over the days during the training, their responses were significantly reduced compared with the controls on day 1 (19.21 ± 3.24 Hz, control, versus 9.74 ± 0.30 Hz, KO; $p < 0.01$), day 2 (36.33 ± 4.39 Hz, control, versus 16.43 ± 4.01 Hz, KO; $p < 0.01$), and day 3 (59.38 ± 3.82 Hz, control, versus 33.88 ± 4.30 Hz, KO; $p < 0.01$) (Figure 4B). These data suggested that while NMDAR1 deletion did not completely prevent DA neurons from developing conditioned responses (bursting) toward reward-predicting cues, it did, however, greatly lower the robustness of the bursting response, a phenomena that we call DA neuron blunting.

Habit Learning, but Not Goal-Directed Learning, Was Impaired in the Operant Appetitive Conditioning

To assess habit learning, we first tested the mice in a lever-pressing operant-conditioning task. In this task an instrumental

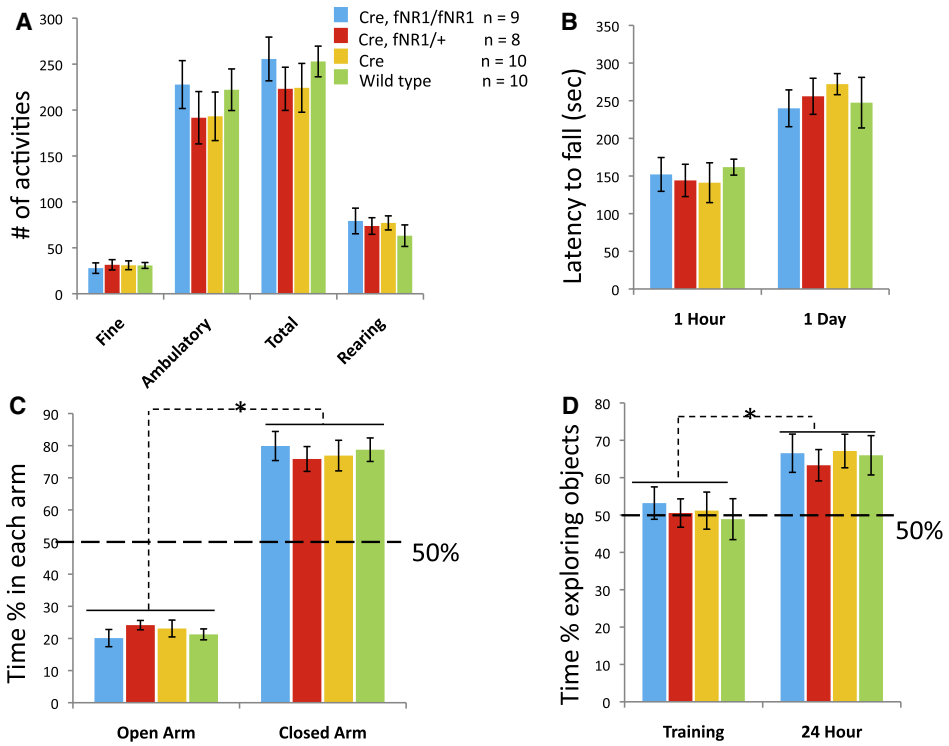


Figure 2. Basic Behavioral Characterization of DA-NR-KO Mice

(A) Locomotor activity tests in a novel environment. Fine movements, ambulatory movements, total movements (total of fine plus ambulatory movements), and rearing were scored. Post hoc comparisons revealed no differences among the mutant and the three control groups.

(B) Rotarod test. Mice were tested 1 and 24 hr after initial training. All mice performed better (as shown by extended latencies) at 24 hr. Post hoc comparisons showed no differences among the mutant and control groups at either time points.

(C) Anxiety level tested using elevated plus maze. Percentages of time spent in each arms were calculated. Post hoc comparisons showed no differences among the genotypes. *$p < 0.05$, Student's t test.

(D) Novel object recognition test. At "24 hour," Time % = time spent exploring novel object/(time spent exploring familiar object + time spent exploring novel object). At "Training," Time % = time spent exploring one object/total time spent exploring both objects. All mice spent more time on the novel object at 24th hr ($p < 0.01$), with post hoc comparisons revealing no differences among the groups tested. Error bars represent SEM. *$p < 0.05$, Student's t test.

action, pressing lever to obtain food, can transform from a goal-directed to a habitual response after extensive training and become progressively less sensitive to devaluation of outcome (Dickinson et al., 1983). The decreased sensitivity can thus be measured as a behavioral readout of habit learning (Figure 5A). Both mutant and control mice learned to press the lever on an extensive training protocol consisting of 4 days of continuous reinforcement (CRF), 2 days of random interval (RI) 30 s, and 6 days of RI 60 s schedules (Dickinson et al., 1983). Mice in both groups increased lever-press rates during the training (CRF days 1–4, RI 30 s days 5 and 6, RI 60 s days 7–12) (Figure 5B). A two-way ANOVA of repeated measures, with days and genotype as factors, showed no effect of genotype ($F_{(1, 231)} = 0.07$), a main effect of days ($F_{(11, 231)} = 51.4$; $p < 0.01$), and no interaction between these factors ($F_{(11, 231)} = 0.269$). This result suggested that the DA-NR1-KO mice have normal wanting of the pellet reward and exhibited normal goal-directed learning.

Lever pressing was then tested after the outcome devaluation. Mice were prefed with either regular mouse chow to which they had been exposed in their regular home cages (nondevalued condition/control), or purified high-energy pellets that are identical to the rewards earned during lever-press sessions (devalued condition). Feeding with mouse chow was used as a control for the overall level of satiety, causing little reduction in the rewarding value of the purified high-energy pellets. Levers were inserted in the 5 min long probe test that immediately followed the 1 hr unlimited food exposure (pellets or chow). No pellets were given during the tests. Comparing numbers of lever press during the tests showed that, while no differences were found between the mutant and the control mice on nondevalued condition ($p = 0.94$) or between the devalued and nondevalued conditions ($p = 0.153$) in the control group, there was a significant difference in the mutant mice between devalued and nondevalued conditions ($p < 0.01$). Furthermore, there was also a significant difference between the mutant and control mice on devalued condition ($p < 0.05$). A two-way ANOVA of repeated measures, with treatment and genotype as factors, showed an interaction between the two factors ($F_{(1, 21)} = 4.98$; $p < 0.05$) (Figure 5C). These suggested that the conditional knockout mice failed to develop the lever-pressing habit despite extensive training, and their action stayed goal directed.

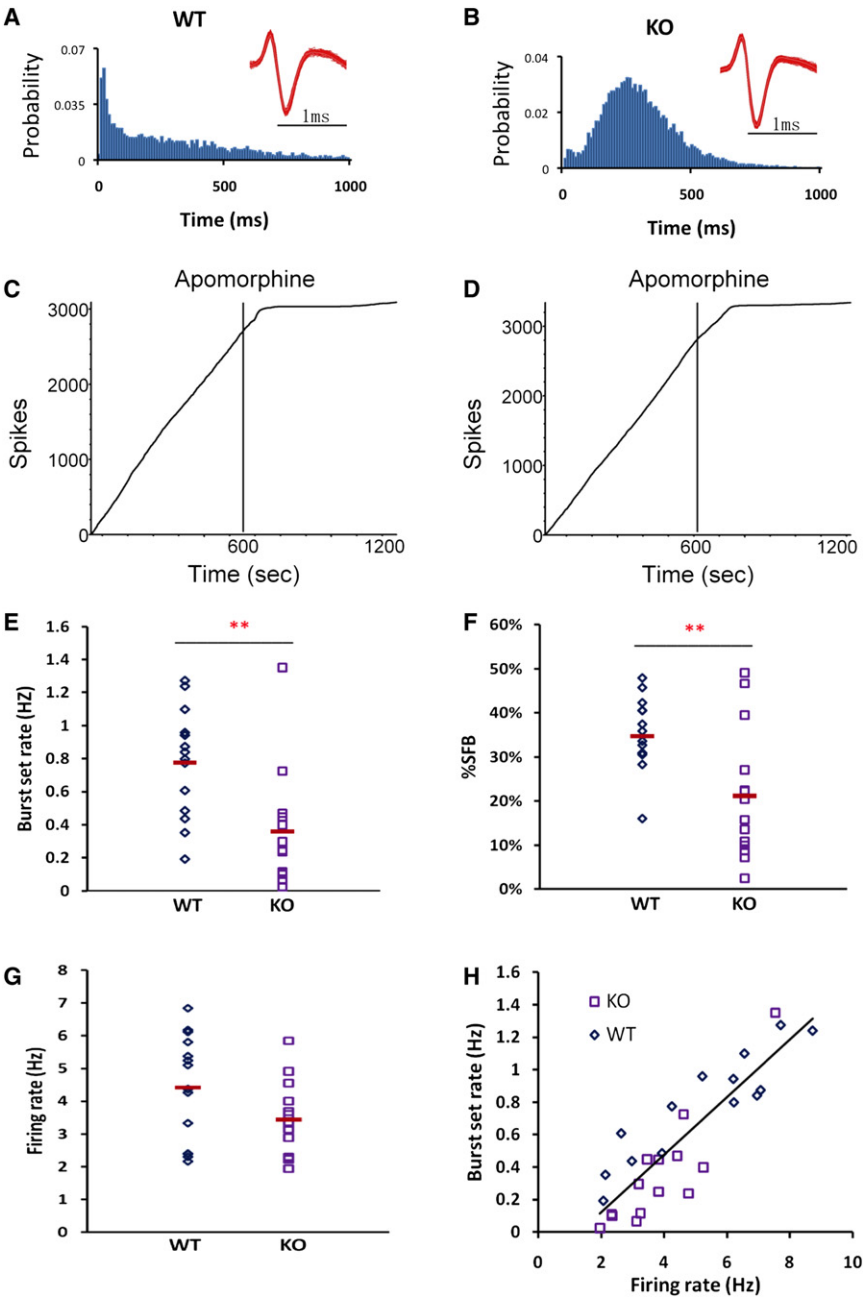


Figure 3. Burst Firing by DA Neurons Is Impaired in KO Mice

(A and B) Sample waveform of a DA neuron recorded from a control (A) and a KO (B) mouse, and corresponding ISI histogram (10 ms bins).

(C and D) Cumulative spike activity of an example of a putative DA neuron from the control (C) and KO (D) mouse in response to the dopamine receptor agonist apomorphine (1 mg/kg, i.p.).

(E) Burst set rate by DA neurons from control and KO mice.

(F) Percent spikes fired in bursts by DA neurons from control and KO mice.

(G) Correlation between burst set rate and firing rate.

(H) Frequency of nonbursting spikes. See text for all the statistics.

Spatial Navigation Habit, but Not Spatial Memory, Was Impaired in the Positively Reinforced Plus Maze

Habit learning was then assessed in a navigation-based paradigm using plus maze place/response-learning tasks (Devan and White, 1999; Packard, 1999; Packard and McGaugh, 1996). Littermates in genotypes *Slc6a3+/Cre; fNR1/+*, *Slc6a3+/Cre*, and wild-type served as three control groups for the DA-NR1-KO mice. The maze was built with transparent walls and placed in a room furnished with spatial cues. The schematic training and testing schedules are shown in Figure 6A. Naive animals, always starting from the same location in the maze (the "south" arm), were trained to find a fixed target site (in the "east" arm) (Training I in Figure 6A). In order to facilitate developing habit-based navigation, the north and the west arms were both closed. It has been shown that under this paradigm, normal mice would learn to search the target using spatial reference memory after moderate training but would switch to habitual navigation after extensive training (Packard and McGaugh, 1996). Probe trials, during which the start location switched from the "south" arm to the "north" arm, were given at different time points to allow dissociation of the spatial and habitual strategies. Thus, mice using the "habit strategy" were predicted to turn right (into the "west" arm), whereas the "spatial" mice, guided by distal spatial cues, were predicted to go to the "east" arm, where the target resided during training.

All mice were trained in ten trials per day for 5 consecutive days before the first probe trial on day 6 (Probe 1 in Figure 6A). During this probe trial the DA-NR1-KO group and control mice showed similar preferences (χ^2 [3, n = 43] = 0.346; p = 0.951) for the "spatial" strategy, opting to turn left toward the "east" arm (Figure 6B), suggesting that they had similarly acquired the spatial memory and that they shared comparable motivation. All mice were then trained for 10 additional days before the second probe trial (Probe 2 in Figure 6A) on day 17. During this probe trial no significant differences were found among the three control groups (χ^2 [2, n = 29] = 0.499; p = 0.779). As a group, control mice opted to "turn right" (and into the "west" arm) significantly more on day 17 than on day 6 (χ^2 [1, n = 29] = 22.587; p = 0.00000201), indicating a learned "habit"-based searching strategy. In contrast, less than 10% of the DA-NR1-KO mice (compared with 80% of control mice) (mutants versus controls: χ^2 = 7.244; p = 0.007)

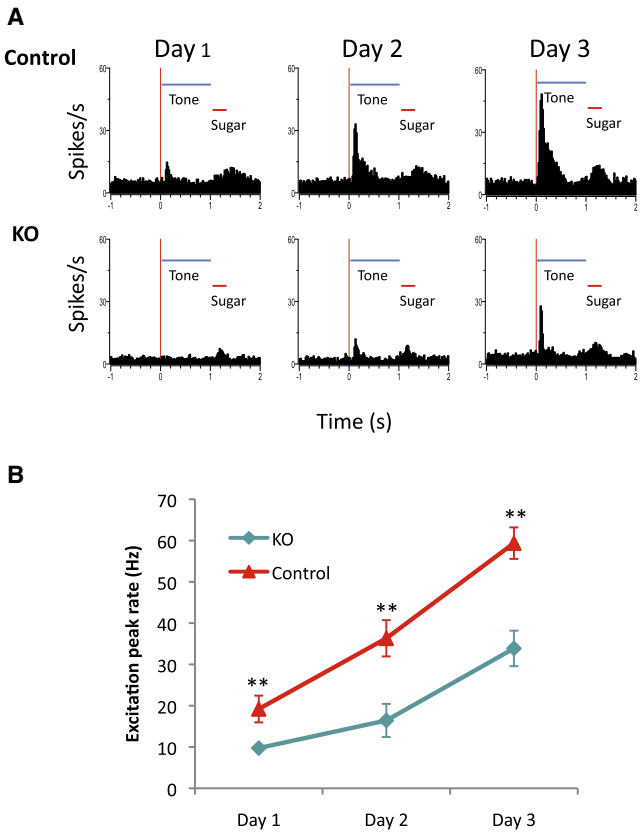


Figure 4. Responses of Putative Dopamine Neurons in 3 Day Reward Test

(A) Peri-event histograms of an example DA neuron recorded from control and KO mouse in response to the same conditioned tone (5 kHz, 1 s) that predicted a sugar pellet delivery in a 3 day session.

(B) Excitation peak firing rate of DA neurons from control (red line, n = 16) and KO (blue line, n = 14) in response to reward during these 3 days. Error bars represent SD. **p < 0.01, excitation peak rate in KO versus control DA neurons, in day 1, 2, and 3.

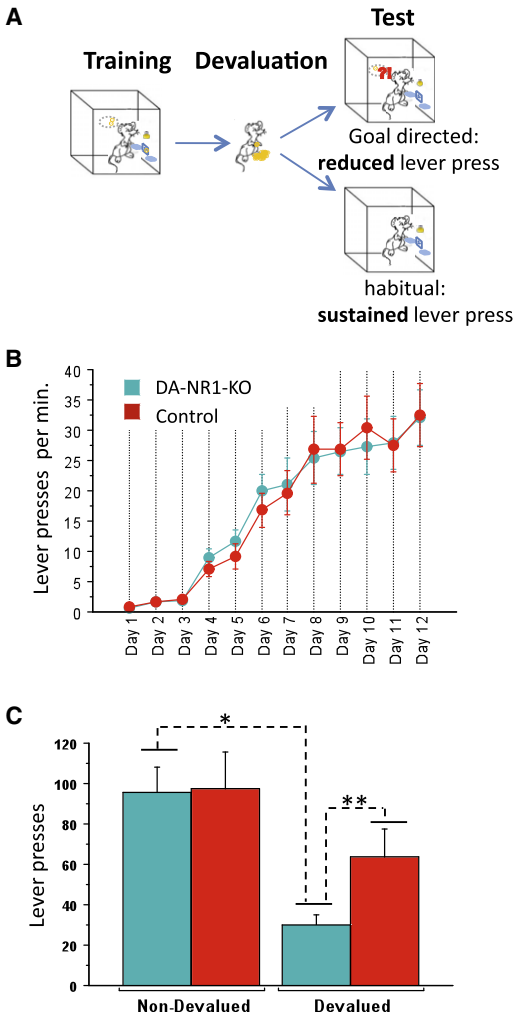


Figure 5. Habit and Goal-Directed Learning Test with Operant Appetitive Conditioning

(A) Paradigm for the training and test.

(B) Lever-pressing speed during training days. No significant difference was discovered at any day of the training between the mutant and control groups.

(C) Lever pressing in the 5 min extinction test after devaluation of outcome and nondevaluation of outcome. Mutant mice showed significantly reduced lever pressing at the devalued condition versus both control mice at devalued condition and mutants at nondevalued condition. No significant reduction was detected for the control mice between the devalued and nondevalued conditions. (See text for all the statistics.) Error bars represent SEM. *p < 0.01, lever presses by the DA-NR1-KO mice during nondevalued versus devalued tests; **p < 0.05, lever presses during devalued test by DA-NR1-KO versus by control.

opted to turn "right" on day 17 (Figure 6B), suggesting that they failed to learn the "habit"-based strategies and, instead, kept using the "spatial" strategy.

To confirm that the deficits in the plus maze tasks were indeed from habit learning, right after the second probe trial, mice were further challenged in a "relearn after 90° rotation" procedure (Training II, Figure 6A), three trials a day for 2 days within the exact same maze and surrounding cues. During the training, both the west and south arms were blocked. The start box was placed in the "east" arm, and the food rewards were in the "north" arm. Mice were tested in a rotation test on day 19, and accuracies to locate the food were scored. Mice started from the "east" arm with all arms open during the test (Figure 6A). For "habit" mice who had learned to "turn right" during previous training sessions (days 1–16), this new learning was simply a retraining, in which the same habit response (turning "right") would lead them to the new food location. However, for the "spatial" mice, switching of target location from the "east" arm to the "north" arm conflicted with the previously learned spatial relationship and, thus, was predicted to inhibit new learning. As in Figure 6C, the mutants showed significantly less success (turning "right" or into the "north" arm) (χ^2 [3, n = 42] = 11.667; p = 0.0006), whereas no difference was found (χ^2 [3, n = 42] = 0.73; p = 0.694) among the three control groups. This supported the notion that mutant mice failed to learn the habit strategy, even after the extensive training.

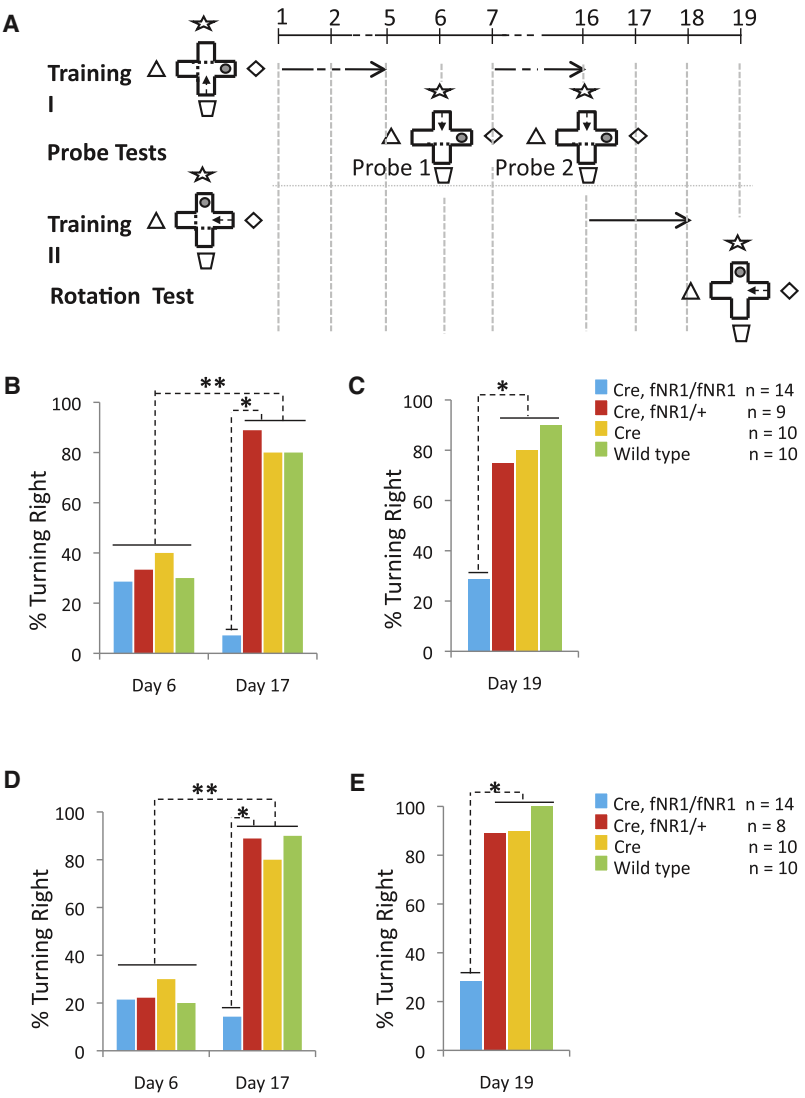


Figure 6. Habit Learning Analyzed Using Plus Maze

(A) Training and testing schedule for both food reward-based and water-based plus maze.

(B and C) Plus maze positively reinforced with food reward.

(D and E) Plus maze negatively reinforced with water.

(B and D) Probe trials on days 6 and 17. Significant differences were found between days 17 and 6 in the control mice. Significant differences were also found on day 17 between the mutant and control groups.

(C and E) Rotation test after 2 days of "relearn after 90° rotation" training. Significant differences were found between the mutant and control groups. (See text for all the statistics.) *$p < 0.01$, mutant versus the controls on day 17 and day 19; **$p < 0.01$, day 17 versus day 6 in the control groups.

Spatial Navigational Habit Learning, but Not Spatial Memory, Was Impaired in the Negatively Reinforced Plus Maze

Because many studies suggested that dopamine is important for reward pathways, we asked whether habit-learning deficits seen in the DA-NR1-KO mice hinged on the nature of the reinforcement. The aforementioned experiments were replicated in a water-based plus maze, in which the sole escape from the water was for mice to locate and climb onto a hidden platform in the end of one arm. This water-based plus maze behavior was driven by the desire to escape from the negative environment and offered an additional opportunity to compare with habit learning based on positive reinforcement such as the seeking of a food reward. All parameters such as maze dimensions, cues used, starting and target locations, number of trials per day, and numbers of days in training remained the same as those in the previous food-rewarded experiments (Figure 6A).

The first probe trial revealed no significant differences between any two of the four genotypes (χ^2 [3, n = 43] = 0.346; p = 0.951). The second probe trial showed that over 80% of the control mice had adopted the "habit" strategy, whereas the mutant mice remained strongly "spatial" (Figure 6D). No differences were found among the three control groups (χ^2 [2, n = 29] = 0.499; p = 0.779). As a group, the control mice opted for the "habit" strategy significantly more on day 17 than on day 6 (χ^2 [1, n = 29] = 22.587; p = 0.00000201). A significantly lower percentage of DA-NR1-KO mice opted to "turn right" (7.14% versus 80% in the control mice; χ^2 [1, n = 43] = 20.904; p = 0.00000483). The deficits in habit learning were further confirmed in the rotation test given after 2 days of the "relearn after 90° rotation" challenge task (Training II, Figure 6A). A significantly smaller proportion of the mutant mice (28.6%) in contrast to 80% of the controls were able to successfully locate the new platform position (one-tailed probability = 0.000388, Fisher's exact test). These data thus agreed with the findings from the above food-rewarded tasks suggesting that the learning deficits were unlikely contingent on the types of reinforcement employed in the training process.

Due to the significant involvement of spatial learning in the plus maze task, mice were tested in a spatial version of the plus maze (Figure 7A). They were trained six trials per day for 6 days to find a hidden platform in the water-filled plus maze. With all four arms open, starting points switched between trials in each day rotating among the distal ends of three arms that did not contain the platform, following a semi-random order. The platform location remained fixed throughout. A probe test was given on day 10, 3 days after the training session ended. During the test, with the platform removed, mice were released to the center of the maze and allowed to search for 60 s. Durations spent by each mouse in each arm were recorded (Figure 7B). Mice from all four groups spent significantly more time searching in the target arm (mutants, $F(3,32) = 101.292$, $p < 0.001$; Cre, fNR1/+, $F(3,28) = 134.996$, $p < 0.001$; Cre, $F(3,36) = 147.806$, $p < 0.001$; wild-type, $F(3, 36) = 294.358$, $p < 0.001$; Newman-Keuls post

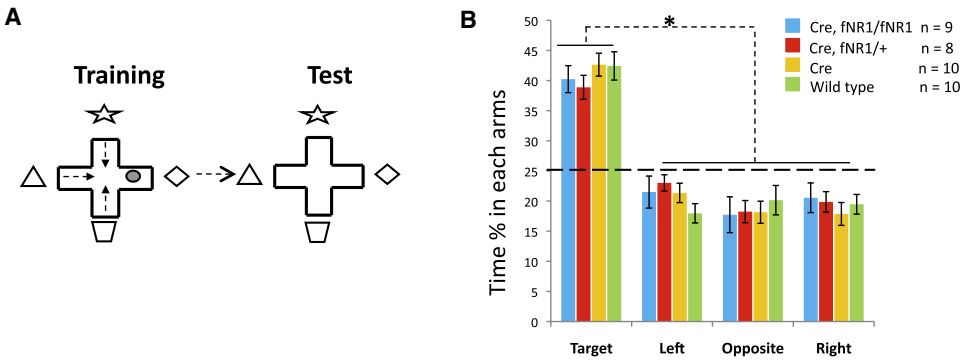


Figure 7. Spatial Memory Test Using Plus Maze

(A) Training and testing paradigms.

(B) Probe trial test. All mice spent more time searching in the target arm. No difference was detected between the mutant and control groups. *$p < 0.01$, time spent in the target arm versus in the other three arms. Error bars represent SEM. (See text for all the statistics.)

hoc comparison [the target arm compared to all the other arms], $p < 0.01$ for all genotypes). No differences were found between the mutant and any control groups, suggesting that spatial learning abilities were unlikely a factor causing the habit-learning deficits observed in the DA-NR1-KO mice.

Habit Learning in a Nonspatial Zigzag Maze-Based Habit Task Was Impaired

Instead of compromising habit learning per se, DA-specific NR1 deletion could have skewed the competition between "spatial" and "habit" memory systems in the plus maze task. In order to investigate this possibility, we designed a nonspatial "zigzag maze" task as a more direct measurement of habit learning. As shown in Figure 8A, the water-filled zigzag maze consisted of eight arms similar in length. Mice were trained to escape onto a hidden platform. Six different starting points were chosen, each paired with its own location of the hidden platform. The platform locations were chosen so that they would be reached after two consecutive right turns from the start point. All mice were trained 12 trials per day for 10 days. To facilitate developing the turning habits, some arms were blocked (red lines) so that mice were only allowed the correct turn at each intersection. A probe test was given on day 11 in which mice were placed at a random start location. Some arms in the maze remained blocked (red lines), but unlike in training, mice were allowed to choose between turning "left" or "right" at two intersections (Figure 8A). Mice were scored for whether they finished the two consecutive right turns (counted as "successful"). No differences were found among the three control genotypes (all between 90% and 100%, χ^2 [2, n = 29] = 1.968; p = 0.374) (Figure 8B), and they were pooled. The conditional knockout mice showed a significantly lower successful rate in making the two consecutive right turns (one-tailed probability = 0.000196, Fisher's exact test), again suggesting that the DA-NR1-KO mice are defective in developing the navigation habit.

DISCUSSION

Here, we studied mutant mice with DA neuron-selective NR1 deletion using a set of behavioral tasks as well as in vivo neural-recording techniques. Behavioral analysis revealed that the DA-NR1-KO mice were impaired in several forms of habit learning. In an operant task where both the mutant and control mice learned a goal-directed action in the initial training, extensive training shifted, but only in the control mice, the learned action from goal directed to habitual. In the mutant mice, this action remained goal directed and, thus, sensitive to reward devaluation. Similarly, in plus maze tasks, whereas both mutants and the controls learned to navigate based on spatial cues in initial training, extensive training shifted navigation from spatial into habitual also only in the controls, while the mutants' navigation remained spatially oriented. Such deficits in habit learning were observed in both positively reinforced and negatively reinforced tasks. This is consistent with our recent recordings showing that DA neurons employ a convergent encoding strategy for processing both positive and negative values (Wang and Tsien, 2011). One notable finding of those in vivo recording experiments was that some DA neurons exhibit a stimulus-suppression-then-rebound-excitation type firing pattern in response to negative experiences (Wang and Tsien, 2011). This offset-rebound excitation may encode information reflecting not only a relief at the termination of such fearful events but, perhaps, provide some sort of motivational signals (e.g., motivation to escape).

Therefore, our data strongly suggested that NMDAR functions in DA neuron be essential for habit learning. A previous study by Zweifel et al. (2009) reported that the DA neuronal-selective NR1 KO mice were impaired in learning a water maze task and also impaired in learning a conditioned response in an appetitive T maze task, seemingly in disagreement with our results of normal spatial learning and goal-directed learning. The experimental conditions used in their studies were, however, quite different from those in ours. The water maze deficit was transient and detectable only during the very early part (day 2 in a 5 day session) of their training sessions. The T maze was a goal-directed paradigm that likely also involved mice learning context association between landmarks and rewards. Additionally, the action-reward contingency was also different than that in the operant paradigm that we used. It is very likely that factors such as task difficulties, amount of training, cue saliencies, temporal and spatial contingencies between the CS, and the

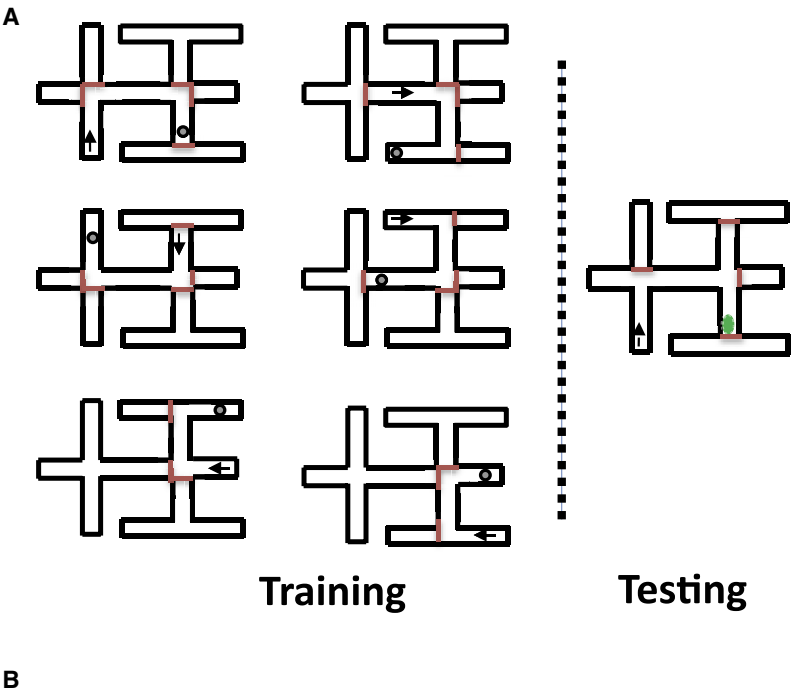


Figure 8. Habit-Learning Test Using Zigzag Maze

(A) Training and testing paradigms. Red lines indicate blocked entrances, arrows starting points, and black circles target locations.

(B) Turning test. Mutant mice showed a significantly reduced rate of success versus the controls. *$p < 0.001$, mutant versus control groups. (See text for all the statistics.)

rewards can affect the type and amount of involvement by DA neurons. Using in vivo neural recordings, we observed that although the response to cue-reward association is much attenuated in DA-NR1-KO neurons in term of both response peak amplitude and duration, these DA neurons, nonetheless, still could form the cue-reward association. Interaction between the blunted responsiveness of DA and test conditions may leave some goal-directed learning impaired by the NR1 deletion, whereas spare some others. A good example is that in a report published later by the same group, the authors reported that the NR1 KO mice were normal in a goal-directed learning paradigm (Parker et al., 2010). In our study the test conditions and amount of trainings we used allowed the controls as well as the mutants to learn goal-directed and spatial learning normally and indistinguishably. After extensive training under these conditions, the mutant mice could not develop the habit learning, whereas the controls clearly did.

Dopamine is an important modulator for the habit learning (Wickens et al., 2007; Yin and Knowlton, 2006). Most of the current understanding of its involvement in the habit learning has so far centered on the downstream pathways and structures such as the dorsal striatum and more recently the PFC (Wickens et al., 2007; Yin and Knowlton, 2006). Our finding here highlighted the importance of glutamatergic modulations of the DA neuron circuitry itself, in this case mediated by NMDARs, and suggested that this upstream pathway should be considered an integral part of the habit-learning networks. With perception of the environmental stimuli likely carried out by glutamatergic signals, it is conceivable that NMDARs in dopaminergic neurons participate in the controlling and fine-tuning of dopaminergic neuron activity patterns during habit formation. An important part of this regulation is perhaps to create the cue-reinforcement association at an appropriate level in terms of response robustness and overall DA neuron network patterns so that DA neurons would respond accordingly to procedures and cues with higher incentive salience. NMDARs are required in mediating synaptic plasticity in glutamatergic synapses onto DA neurons (Bonci and Malenka, 1999). Our results showed that modulation by NMDARs facilitates bursting of DA neurons toward the learned reward-predicting cues. It is conceivable that the function of NMDARs in regulating phasic firing may be closely linked to its roles in regulating synaptic plasticity. In fact, studies have shown that enhanced synaptic strength onto dopamine neurons may act to facilitate their phasic firing (Stuber et al., 2008).

The blunting of the phasic firing of DA neuron in the mutant mice can contribute or even result in the habit-learning deficits. There are several brain regions involved in habit learning that can be affected by this blunting. The most intuitive one is the striatum. Dopamine signaling has been postulated as the mechanism that trains the striatum, which in turn trains the cortex to establish the appropriate sensorimotor associations required for developing habits (Ashby et al., 2010; Wickens et al., 2007).

Dopamine modulates the plasticity in the corticostriatal synapses, facilitating induction of LTP in conditions that would otherwise induce LTD. This facilitation requires dopamine D1 receptor (Calabresi et al., 2000). The low affinity of D1 receptors toward dopamine coupled with the fast dopamine reuptake (Cragg et al., 1997) in the striatum likely makes the dopamine modulation sensitive to the blunting of phasic release. From a network point of view, it has been reported that dopamine can cause changes in the coordinated activity of neuronal ensembles in corticostriatal circuits and by doing so "gate" the inputs in those downstream regions (Costa et al., 2006). Thus, when dopamine level is low, such as when bursting activities are insufficient, it fails to produce and reinforce these networks' connectivity underlying habit formation. Other than the striatum, reduced bursting of DA neurons may also affect activities of structures such as the PFC of which lesion of the medial infralimbic area was reported to impair expression of a learned habit (Coutureau and Killcross, 2003). Studies have shown that tonic dopamine concentration in the prefrontal area, likely due to the relatively slower dopamine reuptake (Seamans and Yang, 2004), may be affected by previous phasic dopamine release (Matsuda et al., 2006). The presence of background dopamine signal converts LTD to potentiation. This "priming" requires time to develop and requires D1 and D2 receptors, both of which have low affinity to dopamine. It is very likely that this phasic release-induced "priming" could also be affected by the amount of DA neurons bursting, thus, by blunting of DA response. It will be of great interest to dissect the various roles of those different brain regions in habit formation in future studies.

It is also important for future research to further analyze the contributions of NMDARs within different dopamine subpopulations, and temporally within different phases of habit learning. The potential subregional circuitry within the DA neuron populations in the VTA and SNr regions can be highly crucial for integrating distinct cortical and subcortical inputs (Grace et al., 2007; Lammel et al., 2011; Lisman and Grace, 2005). Thus, it is conceivable that additional subregional-specific manipulations and analyses could further elucidate how the glutamatergic regulation of DA neurons, as revealed by our current study, modulates habit formation.

In summary our study has provided several important insights about NMDAR in DA neurons and habit learning. First, NMDARs in DA neurons are required for learning habits, including appetitive lever pressing and spatial navigational habits. Second, the dependence of habit learning on NMDARs in DA neurons was observed in both positively and negatively reinforced trainings. Third, DA neurons lacking the NMDARs can still form the cue-reward association but with greatly reduced phasic activity as well as conditioned response robustness. Taken together, our results suggest that the NMDARs in DA neurons are an important modulator of DA neurons' response robustness in cue-reward association and an essential element underpinning habit learning.

EXPERIMENTAL PROCEDURES

Subjects

Mice carrying alleles of *NMDAR1* flanked by loxP sites (*fNR1*) were bred with Slc63a Cre transgenic mice. Offspring were genotyped by PCR for both the Cre transgene and for the floxed *NMDAR1* (*fNR1*) locus. Mice used in these experiments have been bred for at least five generations onto the C57/BL6 background. Animals were maintained on a 12 hr light/dark cycle in the Georgia Health Sciences University animal care facility. Except for when specified in experiment, such as when food pellets were used as rewards, food and water were given ad libitum. All procedures relating to animal care and treatment conform to the Institutional and NIH guidelines. For behavioral tests in the study, we used male mice around 1 year old in age. These animals have been prescreened to make sure that they have normal vision and hearing capacity.

Immunohistochemistry

Mice were perfused transcardially with 4% paraformaldehyde (PFA) in 1× PBS followed by a postfixation in 4% PFA overnight. Coronal sections (50 μm thick) were cut on a vibratome and collected in 0.5% PFA in 1× PBS and stored at 4°C before use. For double-immunofluorescent staining of β-galactosidase and TH, sections were incubated at 4°C overnight with gentle shaking in primary antibody (anti-β-galactosidase [pAb] 1/5,000, Invitrogen; anti-TH [monoclonal antibody] 1/1,000) in a buffer containing 0.05% Tween 20, 10% normal goat serum, and 1× PBS following preincubation in 10% normal goat serum and 1× PBS at room temperature for 2 hr. The sections were then incubated with Alexa-conjugated secondary antibodies (1/200; Invitrogen) at room temperature for 2 hr. β-Galactosidase IR was visualized by Alexa 568 and TH IR by Alexa 488. A similar procedure was employed to double stain NMDAR1 and TH except that anti-NR1 (polyclonal antibody 1:100; Chemicon, Temecula, CA, USA) was used as the primary antibody for NMDAR1. The sections were incubated with Alexa-conjugated secondary antibodies (1/200; Invitrogen) at room temperature for 2 hr. NMDAR1 was visualized by Alexa 488 and TH IR by Alexa 594. Fluorescent images were captured with a confocal laser-scanning microscope and an epifluorescence microscope.

Elevated Plus Maze

This apparatus consists of a center platform (5 × 5 cm) 37 cm off the ground with four branching arms (30 cm long and 5 cm wide). Two of the four arms are open, and the other two arms are enclosed by black walls (20 cm high). Testing was performed during light phase in a dimly lit room (50 lux). Animals were placed on the center platform and scored for arm entries and time spent in each arm. Percentages of time animals spent in the open arms were calculated as the final readout of anxiety. Unpaired t tests were used to compare the significance between the different genotypes.

Rotarod

RotaRod analysis was performed using the mouse version of ROTA-ROD manufactured by San Diego Instruments (San Diego, CA, USA). Mice were trained by allowing them to run on a rotarod rotating at 30 rpm for a total time span of 5 min. (Time counting was stopped when mice dropped until they were put back onto the rotarod again.) During the tests mice were again placed on top of the rotarod, which rotated at 30 rpm. Durations of each mouse that stayed on the rotarod (latency to fall) were recorded. Any mice remaining on the apparatus 300 s after the starts were removed, and the time was scored as 300 s. Unpaired t tests were used to compare the significance between the latencies in different genotypes.

Open-Field Activity Test

Locomotor activity was measured by scoring beam breaks in activity chambers (San Diego Instruments). Prior to open-field tests, animals were handled for 2 consecutive days. Standard rat cages were used as the novel open field for the mice tested. Locomotor activities were recorded for 1 hr and scored for both 5 min and 1 hr. Unpaired t tests were used to compare the significance in fine movements, ambulatory movements, and rearing between the different genotypes.

Instrumental Training

Mice were placed on a food-deprivation schedule to reduce their weight to 80%–85% of their baseline weight. They were fed for 2 hr with mouse chow

in their home cages each day after training. Water was available at all times in the home cages.

Training and testing took place in eight Med Associates operant chambers (21.6 cm length × 17.8 cm width × 12.7 cm height) housed in boxes with sound-attenuating walls. Each chamber was equipped with a food magazine, two retractable levers, one on each side of the magazine, and a 3 W, 24V house light mounted on the same wall, but above the food magazine. Bio-Serv 20 mg pellets from a dispenser into the magazine were used as reward. The software Med-PC-IV from Med Associates was used for equipment control and behavior recording.

Lever-Press Training

At the beginning of each session, the house light was turned on and the lever inserted. At the end of each session, the light was turned off and the lever retracted. Mice were trained in an initial lever-press training consisting of 4 consecutive days of CRF, during which the mice received a pellet for each lever press. A session would end after 60 min or after the mouse had collected 30 rewards, whichever came first. After CRF, mice were trained with RI schedules to generate habitual lever pressing (Dickinson et al., 1983). The training started with 2 days on RI 30 s, with a 0.1 probability of reward availability every 3 s contingent on lever press, and followed by 6 days on the 60 s interval schedule, with a 0.1 probability of reward availability every 6 s contingent on lever pressing. Repeated-measures ANOVA was used to compare lever press between the different genotypes.

Devaluation Tests

A specific satiety procedure was used for outcome devaluation. Mice were given unlimited access within a fixed duration to either the mouse chow to which they had been exposed in their home cages (nondevalued condition/control), or the purified pellets they normally earned during lever-press sessions (devalued condition). The mouse chow served as a control for overall level of satiety. This procedure controls the overall level of satiety and motivational state, while altering the current value of a specific reward. Immediately after 1 hr of unlimited exposure to the pellets or chow, the mice were subjected to a 5 min long probe test. During the probe test the lever was inserted, but no pellet would be delivered in response to lever pressing. This brief extinction test was designed to test whether the acquired lever pressing of the mice was controlled by the action-outcome instrumental contingency or habit (e.g., in response to a antecedent stimuli). On the second day of outcome devaluation, the same procedure was used, except that those animals that received mouse chow on day 1 received pellets on day 2, and vice versa. When grouping, mice were counterbalanced between genotypes and treatment. Repeated-measures ANOVA and unpaired Student's t test were used to compare lever press between the different genotypes as specified in the text.

Plus Maze

The maze consisted of four arms measuring 35 cm long, 6 cm wide, and 35 cm deep, with transparent high walls made of clear Plexiglas. For training positively reinforced with food pellets (20 mg per pellet), animals were maintained at 80%–75% of their free-feeding weight throughout the experiment. For training negatively reinforced with water, water was stained opaque and white with titanium dioxide. A hidden platform was placed 1 inch under the water surface. The training and testing were as described in the text. For plus maze assays, littermates in Slc6a3+/Cre, fNR1/+ (control), Slc6a3+/Cre (Cre control), and wild-type genotypes were chosen as three control groups. Turning of mice in different tests was compared using chi-square tests, as specified in the text, to evaluate the performance of mice from different genotypes. Additionally, repeated-measures ANOVA and unpaired Student's t test, as specified in the text, were used to compare time spent in different arms among mice from the different genotypes.

Zigzag Maze

The shape of the zigzag maze is illustrated in Figure 8A. Each arm measures about 30 cm long, 6 cm wide, and 35 cm deep. The maze was filled with water that was stained opaque and white with titanium dioxide. A hidden platform was placed at a designed location 1 inch under the water surface. Training and tests were done as described in the text. The chi-square test and Fisher's exact test were used to compare the performance of mice from different genotypes.

Surgeries

A 32-channel (a bundle of 8 tetrodes), ultralight (weight <1 g), movable (screw-driven) electrode array was constructed similar to that described previously (Lin et al., 2006; Wang and Tsien, 2011). Each tetrode consisted of four 13 μm diameter Fe-Ni-Cr wires (Stablohm 675, California Fine Wire; with impedances of typically 2–4 MΩ for each wire) or 17 μm diameter Platinum wires (90% Platinum 10% Iridium, California Fine Wire; with impedances of typically 1–2 MΩ for each wire). One week before surgery, mice (3–6 months old) were removed from the standard cage and housed in customized home cages (40 × 20 × 25 cm). On the day of surgery, mice were anesthetized with ketamine/xylazine (80/12 mg/kg, i.p.); the electrode array was then implanted toward the VTA in the right hemisphere (3.4 mm posterior to bregma, 0.5 mm lateral and 3.8–4.0 mm ventral to the brain surface) and secured with dental cement.

Tetrode Recording and Unit Isolation

Two or three days after surgery, electrodes were screened daily for neural activity. If no dopamine neurons were detected, the electrode array was advanced 40–100 μm daily, until we could record from a putative dopamine neuron. Multichannel extracellular recording was similar to that described previously (Lin et al., 2006; Wang and Tsien, 2011). In brief, spikes (filtered at 250–8000 Hz; digitized at 40 kHz) were recorded during the whole experimental process using the Plexon Multichannel Acquisition Processor System. Mice behaviors were simultaneously recorded using the Plexon CinePlex tracking system. Recorded spikes were isolated using the Plexon Offline Sorter software: multiple spike sorting parameters (e.g., principle component analysis, energy analysis) were used for the best isolation of the tetrode-recorded spike waveforms. Combining the stability of multi-tetrode recording and multiple unit-isolation techniques available in Offline Sorter (e.g., principle component analysis, energy analysis), individual VTA neurons can be studied in great detail, in many cases for days.

Reward Conditioning of DA Neurons

Mice were slightly food restricted before reward association training. In reward conditioning, mice were placed in the reward chamber (45 cm in diameter, 40 cm in height). Mice were trained to a tone (5 kHz, 1 s) with subsequent sugar pellet delivery for at least 3 days (40 trials per day; with an interval of 1–2 min between trials). The tone was generated by the A12-33 audio signal generator (5 ms shaped rise and fall; about 80 dB at the center of the chamber) (Coulbourn Instruments). A sugar pellet (14 mg) was delivered by a food dispenser (ENV-203-14P; Med. Associates) and dropped into one of two receptacles (12 × 7 × 3 cm) at the termination of the tone (the other receptacle was used as control, where a sugar pellet was never received).

Analysis of In Vivo Recording Data

Sorted neural spikes were processed and analyzed in NeuroExplorer (Nex Technologies) and MATLAB. Dopamine neurons were classified based on the following three criteria.

(1) Low baseline firing rate (0.5–10 Hz).
(2) Relatively long ISI (all the classified putative dopamine neurons are with ISIs >4 ms within a ≥99.8% confidence level). The shortest ISI we recorded was 4.1 ms under any conditions in our experiment (only well-isolated units with amplitude ≥0.4mV were used for calculation of the shortest ISI). The averaged shortest ISI was 6.8 ± 2.2 ms (mean ± SD; n = 36). In contrast the ISI for nondopamine neurons can be as short as 1.1 ms.
(3) Regular firing pattern when mice were freely behaving (fluctuation <3 Hz). Here, fluctuation represents the SD of the firing rate histogram bar values (bin = 1 s; recorded for at least 600 s).

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.neuron.2011.10.019.

ACKNOWLEDGMENTS

We thank Fengying Huang and Brianna Klein for technical assistance. This work was supported by funds from NIMH, NIA, and Georgia Research Alliance (all to J.Z.T.). L.P.W., F.L., X.S., and J.Z.T. conceived and designed the experiments. L.P.W., F.L., D.W., K.X., and D.H.W. performed the experiments. L.P.W., F.L., D.W., K.X., and J.Z.T. analyzed the data. L.P.W., F.L., X.S., and J.Z.T. contributed reagents/materials/analysis tools. L.P.W., F.L., K.X., and J.Z.T. wrote the paper. All experiments involving animals were performed according to procedures approved by the Georgia Health Sciences University IACUC review board.

Accepted: October 10, 2011
Published: December 21, 2011

REFERENCES

Ashby, F.G., Turner, B.O., and Horvitz, J.C. (2010). Cortical and basal ganglia contributions to habit learning and automaticity. Trends Cogn. Sci. (Regul. Ed.) *14*, 208–215.

Bonci, A., and Malenka, R.C. (1999). Properties and plasticity of excitatory synapses on dopaminergic and GABAergic cells in the ventral tegmental area. J. Neurosci. *19*, 3723–3730.

Calabresi, P., Gubellini, P., Centonze, D., Picconi, B., Bernardi, G., Chergui, K., Svenningsson, P., Fienberg, A.A., and Greengard, P. (2000). Dopamine and cAMP-regulated phosphoprotein 32 kDa controls both striatal long-term depression and long-term potentiation, opposing forms of synaptic plasticity. J. Neurosci. *20*, 8443–8451.

Costa, R.M., Lin, S.C., Sotnikova, T.D., Cyr, M., Gainetdinov, R.R., Caron, M.G., and Nicolelis, M.A. (2006). Rapid alterations in corticostriatal ensemble coordination during acute dopamine-dependent motor dysfunction. Neuron *52*, 359–369.

Coutureau, E., and Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. Behav. Brain Res. *146*, 167–174.

Cragg, S., Rice, M.E., and Greenfield, S.A. (1997). Heterogeneity of electrically evoked dopamine release and reuptake in substantia nigra, ventral tegmental area, and striatum. J. Neurophysiol. *77*, 863–873.

Devan, B.D., and White, N.M. (1999). Parallel information processing in the dorsal striatum: relation to hippocampal function. J. Neurosci. *19*, 2789–2798.

Dickinson, A., Nicholas, D.J., and Adams, C.D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. Q. J. Exp. Psychol. B *35B*, 35–51.

Engblom, D., Bilbao, A., Sanchis-Segura, C., Dahan, L., Perreau-Lenz, S., Balland, B., Parkitna, J.R., Luján, R., Halbout, B., Mameli, M., et al. (2008). Glutamate receptors on dopamine neurons control the persistence of cocaine seeking. Neuron *59*, 497–508.

Fama, R., Sullivan, E.V., Shear, P.K., Stein, M., Yesavage, J.A., Tinklenberg, J.R., and Pfefferbaum, A. (2000). Extent, pattern, and correlates of remote memory impairment in Alzheimer's disease and Parkinson's disease. Neuropsychology *14*, 265–276.

Faure, A., Haberland, U., Condé, F., and El Massioui, N. (2005). Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. J. Neurosci. *25*, 2771–2780.

Faure, A., Leblanc-Veyrac, P., and El Massioui, N. (2010). Dopamine agonists increase perseverative instrumental responses but do not restore habit formation in a rat model of Parkinsonism. Neuroscience *168*, 477–486.

Grace, A.A., Floresco, S.B., Goto, Y., and Lodge, D.J. (2007). Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. Trends Neurosci. *30*, 220–227.

Kauer, J.A., and Malenka, R.C. (2007). Synaptic plasticity and addiction. Nat. Rev. Neurosci. *8*, 844–858.

Knowlton, B.J., Mangels, J.A., and Squire, L.R. (1996). A neostriatal habit learning system in humans. Science *273*, 1399–1402.

Lammel, S., Ion, D.I., Roeper, J., and Malenka, R.C. (2011). Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. Neuron *70*, 855–862.

Lin, L., Chen, G., Xie, K., Zaia, K.A., Zhang, S., and Tsien, J.Z. (2006). Large-scale neural ensemble recording in the brains of freely behaving mice. J. Neurosci. Methods *155*, 28–38.

Lisman, J.E., and Grace, A.A. (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron *46*, 703–713.

Matsuda, Y., Marzo, A., and Otani, S. (2006). The presence of background dopamine signal converts long-term synaptic depression to potentiation in rat prefrontal cortex. J. Neurosci. *26*, 4803–4810.

Overton, P., and Clark, D. (1992). Iontophoretically administered drugs acting at the N-methyl-D-aspartate receptor modulate burst firing in A9 dopamine neurons in the rat. Synapse *10*, 131–140.

Overton, P.G., Richards, C.D., Berry, M.S., and Clark, D. (1999). Long-term potentiation at excitatory amino acid synapses on midbrain dopamine neurons. Neuroreport *10*, 221–226.

Packard, M.G. (1999). Glutamate infused posttraining into the hippocampus or caudate-putamen differentially strengthens place and response learning. Proc. Natl. Acad. Sci. USA *96*, 12881–12886.

Packard, M.G., and McGaugh, J.L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. Neurobiol. Learn. Mem. *65*, 65–72.

Packard, M.G., Hirsh, R., and White, N.M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: evidence for multiple memory systems. J. Neurosci. *9*, 1465–1472.

Parker, J.G., Zweifel, L.S., Clark, J.J., Evans, S.B., Phillips, P.E., and Palmiter, R.D. (2010). Absence of NMDA receptors in dopamine neurons attenuates dopamine release but not conditioned approach during Pavlovian conditioning. Proc. Natl. Acad. Sci. USA *107*, 13491–13496.

Saal, D., Dong, Y., Bonci, A., and Malenka, R.C. (2003). Drugs of abuse and stress trigger a common synaptic adaptation in dopamine neurons. Neuron *37*, 577–582.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. J. Neurophysiol. *80*, 1–27.

Seamans, J.K., and Yang, C.R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. Prog. Neurobiol. *74*, 1–58.

Stuber, G.D., Klanker, M., de Ridder, B., Bowers, M.S., Joosten, R.N., Feenstra, M.G., and Bonci, A. (2008). Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. Science *321*, 1690–1692.

Tsien, J.Z., Huerta, P.T., and Tonegawa, S. (1996). The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. Cell *87*, 1327–1338.

Ungless, M.A., Whistler, J.L., Malenka, R.C., and Bonci, A. (2001). Single cocaine exposure in vivo induces long-term potentiation in dopamine neurons. Nature *411*, 583–587.

Wang, D.V., and Tsien, J.Z. (2011). Convergent processing of both positive and negative motivational signals by the VTA dopamine neuronal populations. PLoS One *6*, e17047.

Wang, L.P., Li, F., Shen, X., and Tsien, J.Z. (2010). Conditional knockout of NMDA receptors in dopamine neurons prevents nicotine-conditioned place preference. PLoS One *5*, e8616.

Wickens, J.R., Horvitz, J.C., Costa, R.M., and Killcross, S. (2007). Dopaminergic mechanisms in actions and habits. J. Neurosci. *27*, 8181–8183.

Yin, H.H., and Knowlton, B.J. (2006). The role of the basal ganglia in habit formation. Nat. Rev. Neurosci. *7*, 464–476.

Zhuang, X., Masson, J., Gingrich, J.A., Rayport, S., and Hen, R. (2005). Targeted gene expression in dopamine and serotonin neurons of the mouse brain. J. Neurosci. Methods *143*, 27–32.

Zweifel, L.S., Argilli, E., Bonci, A., and Palmiter, R.D. (2008). Role of NMDA receptors in dopamine neurons for plasticity and addictive behaviors. Neuron *59*, 486–496.

Zweifel, L.S., Parker, J.G., Lobb, C.J., Rainwater, A., Wall, V.Z., Fadok, J.P., Darvas, M., Kim, M.J., Mizumori, S.J., Paladini, C.A., et al. (2009). Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. Proc. Natl. Acad. Sci. USA *106*, 7281–7288.

AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 145:382–389 (2011)

Novelty-Seeking DRD4 Polymorphisms are Associated with Human Migration Distance Out-of-Africa After Controlling for Neutral Population Gene Structure

Luke J. Matthews[1]* and Paul M. Butler[2]

[1]*Department of Human Evolutionary Biology, Harvard University, 11 Divinity Ave., Cambridge, MA 02138*
[2]*Department of Neurology, Boston University School of Medicine, 72 East Concord St., 5th Floor, B528, Boston, MA 02118*

KEY WORDS functional polymorphisms; novelty-seeking trait (NS), natural selection; cultural neuroscience; human variation

ABSTRACT Numerous lines of evidence suggest that *Homo sapiens* evolved as a distinct species in Africa by 150,000 years before the present (BP) and began major migrations out-of-Africa ~50,000 BP. By 20,000 BP, our species had effectively colonized the entire Old World, and by 12,000 BP *H. sapiens* had a global distribution. We propose that this rapid migration into new habitats selected for individuals with low reactivity to novel stressors. Certain dopamine receptor D4 (DRD4) polymorphisms are associated with low neuronal reactivity and increased exploratory behavior, novelty seeking, and risk taking, collectively considered novelty-seeking trait (NS). One previous report (Chen et al.: Evol Hum Behav 20 (1999) 309–324) demonstrated a correlation between migratory distance and the seven-repeat (7R) VNTR DRD4 allele at exon 3 for human populations. This study, however, failed to account for neutral genetic processes (drift and admixture) that might create such a correlation in the absence of natural selection. Furthermore, additional loci surrounding DRD4 are now recognized to influence NS. Herein we account for neutral genetic structure by modeling the nonindependence of neutral allele frequencies between human populations. We retest the DRD4 exon 3 alleles, and also test two other loci near DRD4 that are associated with NS. We conclude there is an association between migratory distance and DRD4 exon 3 2R and 7R alleles that cannot be accounted for by neutral genetic processes alone. Am J Phys Anthropol 145:382–389, 2011. © 2011 Wiley-Liss, Inc.

Dopamine neurotransmission is fundamental to human appetitive and reward-seeking behaviors, movement, cognition, and sociability. DRD4 is a metabotropic inhibitory receptor, which acts to decrease intracellular adenylyl cyclase production following dopamine binding (Van Tol et al., 1991; Oldenhof et al., 1998). Neuroanatomically, DRD4 is differentially expressed, with the prefrontal cortex containing the highest density of receptors that act to inhibit downstream neural firing (Dulawa et al., 1999).

Humans exhibit several DRD4 polymorphisms, which include a well-described 48 base-pair (bp) variable number tandem repeat (VNTR) in exon 3 (Ding et al., 2002; Grady et al., 2003). The most common VNTR alleles are 2-, 4-, and 7-repeats (2R, 4R, 7R), which together account for over 90% of all variation at this locus (Chang et al., 1996; Ding et al., 2002). Across all human populations 4R allele frequencies are highest, 2R is common in Asians but rare in Native Americans, and 7R is common in the Americans but rare in Asians and Africans (Chang et al., 1996; Wang et al., 2004). 7R and 2R alleles are associated with a partial loss of DRD4-mediated prefrontal inhibition because of blunted second messenger response. Compared with the 4R, the 2R and 7R alleles result in 40 and 80% reduction in intracellular second messenger response, respectively (Asghari et al., 1995; Wang et al., 2004).

Promoter polymorphisms, including a −521 C/T single nucleotide polymorphism (SNP) and a 120-bp tandem duplication located 1.2 kilo-bp upstream of the initiation codon, are associated with similar neurophysiologic downstream effects as 2R and 7R exon 3 VNTR alleles (Kamakura et al., 1997; Seaman et al., 1999; Lowe et al., 2004). There is a well-supported correlation between the DRD4 7R, 2R, 120-bp promoter duplication, and −521 C/T SNP and the personality trait *novelty seeking* (NS) (Tomitaka et al., 1999; Tsai et al., 2004; Reist et al., 2007; Munafò et al., 2008). High NS individuals are considered exploratory, impulsive, excitable, quick-tempered, and extravagant, whereas low NS individuals tend to be rigid, prudent, stoic, reflective, staid, and slow-tempered (Cloninger et al., 1991; Ebstein et al., 1996).

Chen et al. (1999) suggested that global population patterning of 4R and 7R frequency variation can be explained by increased migratory distances of NS individuals (7R variant) compared with that of low NS individuals (4R variant). They proposed NS individuals represented a behavioral phenotype that functioned better under the novel ecological and social circumstances imposed during migration such that migration effectively

Grant sponsors: Department of Neurology and Program in Behavioral Neuroscience, Boston University School of Medicine, Department of Human Evolutionary Biology, Harvard University.

*Correspondence to: Luke Matthews, Department of Human Evolutionary Biology, Harvard University, 11 Divinity Ave., Cambridge, MA 02138. E-mail: ljmatth@fas.harvard.edu

Received 27 August 2010; accepted 11 January 2011

DOI 10.1002/ajpa.21507
Published online 5 April 2011 in Wiley Online Library (wileyonlinelibrary.com).

© 2011 WILEY-LISS, INC.

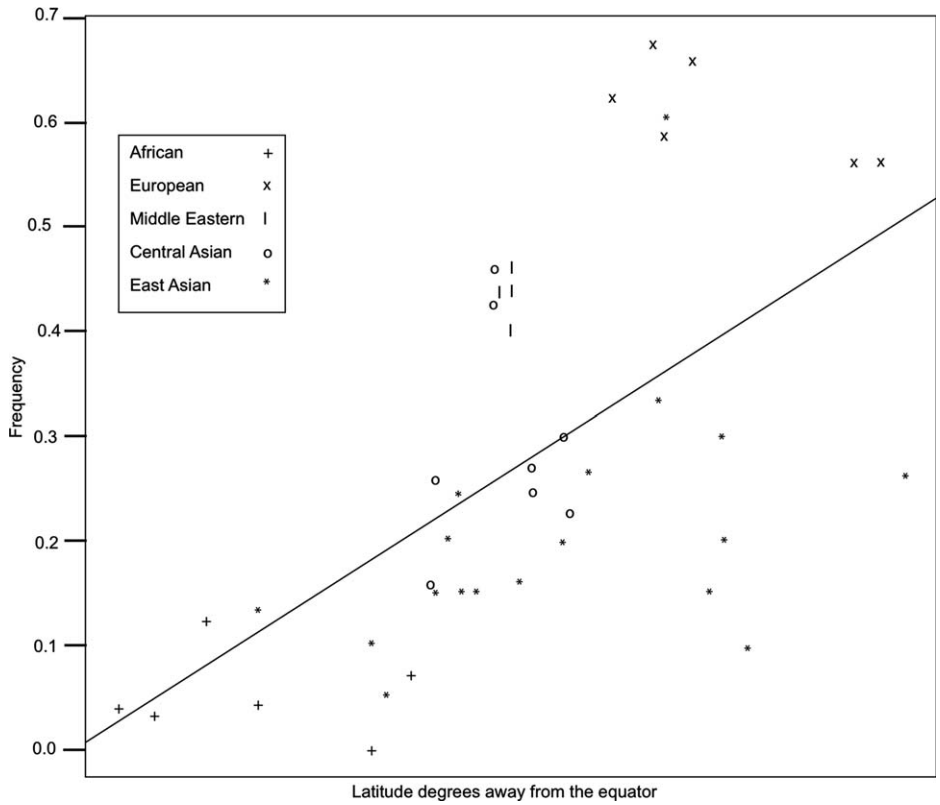


Fig. 1. (Redrawn from Coop et al., Genetics 2010, 185:1411–1423) The effect of statistical nonindependence of population genetic data. The frequencies shown are for a single SNP allele, AGT M235T, in various populations plotted against latitude. Note how populations from the same region frequently cluster on the same side of the regression line, which indicates their statistical nonindependence. Independent data points would scatter randomly across the regression line. The situation is highly analogous to the more well-characterized nonindependence that arises at the species level in a phylogenetic context (Rendall and Di Fiore, 1995; Rohlf, 2006).

selected for NS alleles. They found no association between 7R frequencies and recent migration to the United States by Europeans and East Asians, suggesting that 7R carriers are *not prone to migrate*, but that they may experience higher fitness in response to many generations of migration. This selection hypothesis proposes that migration presents novel ecological and social circumstances which NS individuals deal with more effectively, with less stress response, and thereby experience greater fitness relative to low NS individuals.

Alternatively, if decisions by individuals in recent migrations are fundamentally different from those in ancient migrations, then it could be that NS individuals during ancient migrations were actually more prone to migrate. This nonselection hypothesis still specifies a correlation between migration distance and NS alleles, but the correlation comes about through genetically based tendencies in the decisions of NS individuals.

In this study, we do not attempt to distinguish between these alternative processes (selection or biased decisions based on genotype) that would result in a correlation of NS alleles and migratory distance. Instead, we seek to determine if the correlation itself can be accounted for by neutral genetic processes alone; that is, is it necessary to invoke one of the above non-neutral processes patterning variation in NS alleles? We retested Chen et al.'s (1999) findings for the 7R allele in conjunction with three other alleles (exon 3 VNTR 2R allele, −521 C/T SNP, 120-bp duplication), while controlling for the effects of population history and admixture.

We controlled for neutral genetic processes by using the results of a major recent global genetic survey (Tishkoff et al., 2009) to construct a correlation structure that expressed the correlations expected to arise by drift and admixture in gene frequencies among human populations. Because the out-of-Africa migration proceeded though a process of serial colonization, significant population size bottlenecks have occurred throughout the history of human migration across the globe (Ramachandran et al., 2005; Li et al., 2008; DeGiorgio et al., 2009; Deshpande et al., 2009; Hunley et al., 2009). Thus, observed associations of migratory distance and particular allelic frequencies may result simply from cumulative effects of genetic drift that skew allele frequencies more dramatically with increasing distance from Africa (Klopfstein et al., 2006). This problem exists for any association of a particular allele frequency and an ecological trait, not just migratory distance (Coop et al., 2010; Hancock et al., 2010). In particular, Coop et al., (2010) have proposed that because drift does not systematically affect the frequency of any particular allele, human populations out of Africa only show shared patterns of allelic differentiation because they share the same stochastic drift events. This renders human populations as statistically nonindependent (see Fig. 1). We return to this point in more detail in the Discussion. As pursued by other authors (Coop et al., 2010; Hancock et al., 2010) we attempted to control for neutral genetic processes by attacking the fundamental problem of data point nonindependence with a correlation structure in a generalized

linear model. The correlation structure expressed the expected correlations among population data points under a model of neutral genetic change, and transformed the DRD4 data to remove neutral genetic effects through a likelihood-based linear regression model. We assessed the performance of our correlation structure with a set of previously published neutral loci.

We also used the same neutral loci in an alternative test of the specific association of migratory distance and NS alleles by assessing whether the residual variation of migratory distance regressed on DRD4 alleles was less than exhibited by similar regressions with neutral genetic loci. We asked if frequencies of NS alleles at DRD4 showed a tighter increase with migratory distance than we would expect by chance, chance being reflected in this case by neutrally evolving loci.

MATERIALS AND METHODS

Functional gene data

We searched the published literature for loci surrounding DRD4 with well-characterized in vitro effects on expression or binding, well-characterized effects on personality traits, and with population level sampling from groups whose neutral genetic structure was known. We found 3 loci that were available across 15 or more human populations (Chang et al., 1996; Chen et al., 1999; Osier et al., 2002; Rajeevan et al., 2003; ALFRED, http://alfred.med.yale.edu). These loci were DRD4 exon 3 VNTRs, −521 C/T SNP, and the promoter region 120-bp duplication.

Migratory distance

We derived population locations from the midpoint of the geographic range reported in ALFRED (http://alfred.med.yale.edu; midpoint calculations–http://www.movable-type.co.uk/scripts/latlong.html). We calculated the migratory distance of each population in our study as the great circle distances connecting them across the globe and along known routes of human migration out-of-Africa (Prugnolle et al., 2005; Liu et al., 2006, Mellars, 2006; Long et al., 2009; calculations performed in Google Earth). We set the San as migratory point zero because they were the most anciently diverged population of those included in our study and because they are genetically close to the origin within Africa of human genetic diversity as inferred through coalescent models (Tishkoff et al., 2009). Thus, within Africa we constructed the following routes (San → Mbuti, San → Bantu → Hausa/Biaka). All routes out of Africa went through the Mbuti in central Africa, and then the Druze in the Levant. European and Near Eastern routes led from the Druze directly to Sardinian and Adygei. We constructed a southern route through Asia as Druze → Cambodian → Papuan → Melanesian. A more Northern route through Asia led to the Han followed by the Japanese. The route to the Americans went through the Han and then direct to the Yakut, followed by the New World populations in a North to South sequence. These migration values thus attempt to capture the minimum migratory distance of populations from the origins of our species to their present locations, relative to the other populations in the study. They may underestimate the actual distances that populations have migrated since their inception as distinct demes.

Neutral genetic structure

We controlled for neutral genetic processes of drift, admixture, and the history of population separations by incorporating published D^2 genetic distances into a generalized least squares linear model (see Comparative Analyses below). The D^2 expresses the expected allele frequency differentiation between two populations under an evolutionary model with no selection or mutation (Reynolds et al., 1983). We obtained distances from a recent global study of human genetic differentiation that used data from 1327 polymorphic and neutrally evolving loci from 3945 individuals (Tishkoff et al., 2009).

Comparative analyses

We constructed a generalized least squares (gls) statistical model with migratory distance as the dependent variable and with three independent variables (implemented in R, package "nlme", Pinheiro et al., 2009). One independent variable was the summed frequency for DRD4 2R and 7R alleles, because higher frequencies of *either* allele should produce similar phenotypic effects in the form of increased NS behavior. The other two independent variables were the frequencies of the T SNP at position −521, and the frequencies of the promoter region 120-bp duplication. All independent variables were coded as fixed effects, with the prediction that higher frequencies of each would be associated with greater migratory distance. Because the variance distribution of frequency data rarely conform to assumptions of linear modeling, we arcsine transformed the raw frequency data for subsequent analysis (Sokal and Rohlf, 1995, pp 419–422).

We included a correlation structure in our model to control for the nonindependence of data points in this study (Dow et al., 1984; Pinheiro and Bates, 2000; Rohlf, 2006; Dow, 2007; Coop et al., 2010; Hancock et al. 2010). The allele frequencies of human populations are rendered statistically nonindependent because they share complex histories of drift and admixture that stochastically alter the frequencies of all alleles, including those that are selectively neutral. Chen et al., (1999) recognized this nonindependence, but dealt with it insufficiently by simply scoring whether pairs of populations participated in the same migration routes. Here, we account for the nonindependence with robust measures of genetic distance between populations from neutrally evolving loci. We converted the D^2 distances into expected correlations within the study sample by applying the linear transformation:

$$(\max(D^2) - D^2_{ij}) / \max(D^2)$$

where $\max(D^2)$ is the maximum distance in the matrix and D^2_{ij} is the observed distance between a given pair *i, j*. This transformation produced a matrix of expected correlations among data points ranging from 0 for the two most divergent populations to 1 for each population to itself. This correlation matrix is thus directly analogous in its range of expected correlations to the matrices used in phylogenetic generalized least squares analyses between species (Pagel, 1999; Garland and Ives, 2000; Rohlf, 2006; Coop et al., 2010), but it does not presume bifurcating relationships among the study taxa.

Using our correlation structure to control for neutral population genetic processes, we constructed least squares regression models from each possible combination of the three independent variables, and used Akaike Information

TABLE 1. Allele frequencies as percentages and migratory distance out of Africa

Population	exon 3 R2 + R7[a]	120 duplication[b]	T SNP[b]	Migratory Distance (km)	N-exon 3	N - 120 dup. & T SNP
Adygei	17	74	57	7524	104	108
Bantu	24	*NA*	*NA*	2431	80	*NA*
Biaka	14	41	66	2692	124	140
Cambodian	32	87	74	13540	50	46
Columbian	62	*NA*	*NA*	29997	26	*NA*
Druze	12	69	60	6235	50	140
Han	18	68	66	13123	182	94
Hausa	*NA*	75	67	3792	*NA*	72
Japanese	18	71	53	15774	102	100
Karitiana	60	79	95	31751	108	102
Maya	40	52	86	27593	100	102
Mbuti	16	46	53	2627	74	68
Melanesian	30	43	57	19602	46	44
Papuan	25	*NA*	*NA*	17682	40	*NA*
Pima	23	28	94	24824	70	100
San	0	*NA*	*NA*	0	44	*NA*
Sardinian	15	*NA*	*NA*	8746	26	*NA*
Surui	83	70	89	32124	90	92
Yakut	10	59	70	17728	78	102
Yoruba	*NA*	60	61	3950	*NA*	152

NA = missing data point.
[a] From Chang et al., 1996.
[b] From ALFRED. Sample size, N, given as total chromosomes sampled (2X individuals).

Criterion (AIC) and likelihood ratio tests (LRT) to select the best model. We then assessed the significance of the remaining independent variables through *t*-tests of the slope residuals.

To assess how well our D^2-derived correlation matrix controlled for neutral genetic processes in our final model, we also ran our model with a subset of the neutral alleles used by Tishkoff et al., (2009) to calculate genetic distances. We used the 405 microsatellite and duplication loci provided online by Tishkoff et al., (2009) that were variable for the populations in our study. The full data set used by Tishkoff et al., (2009) for the D^2 calculation includes all alleles at 1327 loci. For each of the 405 loci, we selected one allele at random from individuals in the populations of our final model. We calculated the frequency of these random alleles within each population, applied the arcsine transform to the frequencies, and then ran the same gls analyses for each allele both with and without our correlation structure. When using our correlation structure to control for neutral genetic processes, ideally only 5% of these neutral loci would show a significant association with migratory distance at the 0.05 nominal level for significance. Because the human populations are nonindependent data points, we expected much higher rates of type 1 error (false positives) when the correlation structure was not used.

We also conducted an alternative test of NS alleles that were deemed significant in our linear model with a correlation structure. In this test, we ran a standard least squares regression of migratory distance on the NS allele frequency. If the observed association in fact arose by neutral evolutionary processes, then the goodness of fit between increases in migratory distance and NS allele frequencies should be no stronger than that observed between migratory distance and each of the 405 neutral loci. One standard measure of goodness of fit in a regression model is the residual variation from the best fit regression line. We calculated the residual variation for the best fit model of each of the 405 neutral loci and migratory distance. We subtracted the absolute residual deviations of each of the 405 neutral loci from the absolute residuals from the NS allele frequencies. If the association of NS allele frequencies and migratory distance were no different than that expected under neutral evolution, then on average about half of the neutral loci would have larger residuals and half would have smaller residuals than from the NS allele frequencies. If the association of NS allele frequencies and migratory distance is more than expected to arise from neutral processes, then the difference in residuals should be significantly negative, as the NS residuals should be much smaller than typical of a neutral allele—indicative of a tighter than expected association with migratory distance. We tested whether the calculated differences in residuals were significantly negative with a one-sample Wilcoxon signed rank test.

RESULTS

We assessed which independent variables should be included in the model for migratory distance through a likelihood analysis of all the data points without any missing values (Table 1). The best-supported model to explain out-of-Africa migratory distance included the composite 2R/7R DRD4 allele frequencies (Table 2). The model with this single predictor variable had a better AIC value than all other models. The next most complex models that additionally included either the 120-bp duplication, or the −521 T SNP (see Table 2), did not sufficiently improve the likelihood of the data under the likelihood ratio test ($P = 0.193$). Within the most complex model that included all three predictor variables, variance inflation factors for colinearity among the independent variables were low (2R/7R = 1.4, 120-bp duplication = 1.3, −521 T SNP = 1.3; R package "Design", Harrel, 2009). This indicates that allele frequencies at the loci are largely independent of one another and the alleles at alternative loci can be regarded as unlinked.

The *t*-test of slope residuals in our best model showed a significant positive relationship between the DRD4 2 + 7 repeat elements and migratory distance ($\beta = 18687$, $P = 0.018$, two-tailed, 16 df; Fig. 2). Thus, the summed frequency of 2 and 7 repeat element alleles was more associated with migratory distance than expected by chance.

TABLE 2. Models ordered by AIC value

Model	AIC	log Likelihood
r	265.0	−129.5
r,d	265.3	−128.7
r,s	265.3	−128.7
r,s,d	266.3	−128.1
s	266.6	−130.3
d,s	268.6	−130.3
d	269.7	−131.9

r = DRD4 2 + 7 repeat frequency; d = 120-bp duplication frequency; s = −521 T SNP frequency.

Our tests of 405 neutral markers from Tishkoff et al., (2009) revealed that Type I error rates (false positives) were somewhat elevated under the assumption that the these markers were neutral with respect to selection pressures correlated with migratory distance (observed error = 0.12, expected error = 0.05, z-test of proportions, $P < 0.001$). However, the slope estimates for the neutral loci were not biased up or down, rather they were centered on zero (mean $\beta = 99.3$, t-test for difference from 0, $P = 0.833$, two-tailed). The correlation structure, however, greatly improved the error rate when compared with gls models of the neutral markers that lacked any correction for the nonindependence of the data points (observed error = 0.457). Although the mean of slope estimates from these reduced gls models was more extreme (mean $\beta = -241$) than in the models that included a correlation structure, the estimated slopes again were not significantly different from zero ($P = 0.84$, two-tailed t-test). This can occur because the variance of the slopes across the models without a correlation structure ($\sigma^2 = 6.05 \times 10^8$) was an order of magnitude greater than the variance from slope estimates that used a correlation structure ($\sigma^2 = 8.94 \times 10^7$, Fig. 3).

Given this evidence for slightly inflated Type I error (0.118 instead of 0.05), we corrected the P-value for the significant 2R/7R result based on the observed inflation of error in the neutral loci (a factor of 2.36). The 2R/7R result remained significant after this correction (corrected $P = 0.042$, two-tailed).

The residual variation migratory distance and the DRD4 2 + 7 allele frequency was significantly less than that observed at the 405 neutral loci (median difference in NS and neutral residuals = −3040 km, Wilcoxon signed rank test, $P < 0.001$)

DISCUSSION

This study supports that higher frequencies of novelty seeking (NS) 2 and 7 R alleles at DRD4 exon 3 are associated with population histories of migration, which confirms previous findings specific to the 7R alleles (Chen et al., 1999). Because the out-of-Africa migration proceeded though a process of serial colonization, significant and demographically shared population size bottlenecks have occurred throughout the history of human migration across the globe (Ramachandran et al., 2005; Li et al., 2008; DeGiorgio et al., 2009; Deshpande et al., 2009; Hunley et al., 2009). Thus, observed associations of migratory distance and allelic frequencies may result simply from cumulative effects of genetic drift that skew allele frequencies more dramatically with increasing distance from Africa (Klopfstein et al., 2006). Our study is the first to demonstrate that neutral genetic processes of drift and/or admixture, which produce statistical nonindependence among population data points, cannot explain the association between higher frequencies of NS alleles in DRD4 exon 3 and migration distance.

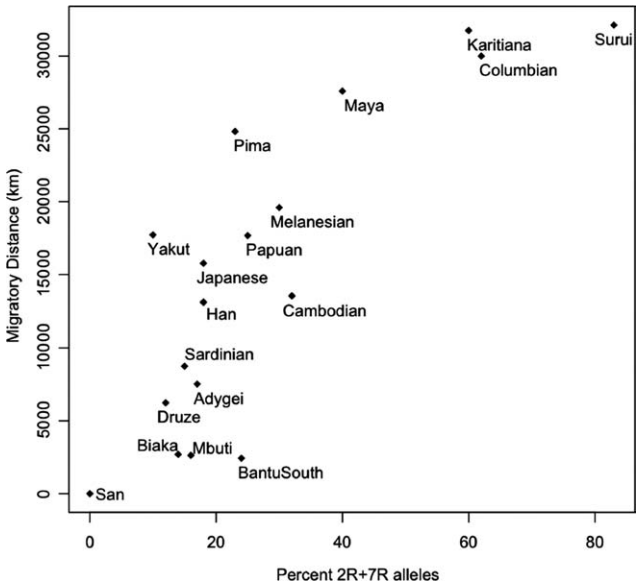


Fig. 2. Migratory distance out-of-Africa versus summed 2R and 7R DRD4 exon 3 alleles.

The gls statistical model with a correlation structure to remove the effects of shared genetic ancestry, admixture, and drift was largely successful. The low variance inflation factors indicate that, with respect to the association with migratory distance, the loci we tested can be regarded as statistically unlinked. Although the correlation structure we used did not reduce the rate of Type I errors (false positive results) to their expected value of 0.05, it did substantially improve the Type I error rate. This marked improvement of the error rate strongly suggests, as proposed recently by other researchers, that statistical associations of allelic frequencies and ecological traits caused by neutral evolution arise primarily from the statistical nonindependence of human populations as data points (Coop et al., 2010; Hancock et al., 2010). Although derived alleles may show a tendency for higher frequency outside Africa as a result of range expansion (Klopfstein et al., 2006), the association in a linear model of the frequencies of neutral alleles and distance from Africa is not biased, as is reflected by the slopes of the neutral loci being centered on zero. As shown in Figure 3, even when we did not control for neutral population structure, the estimated slopes were centered on zero. The fact that using the correlation structure reduced the slope variance while keeping it centered on zero shows that the statistical nonindependence of intraspecific populations behaves much the same as the nonindependence of interspecific datapoints in a phylogenetic context (Martins and Garland, 1991; Rendall and Di Fiore, 1995; Garland et al., 2005; Rohlf, 2006; see also Coop et al., 2010). Phylogenetic nonindependence is well-known to increase the variance of slope estimates and thus inflate Type I error, but it does not systematically bias these estimates; they remain centered on zero (Rohlf, 2006).

This being said, our correlation structure did not reduce the Type I error to its expected level. This can occur if the correlation structure is somewhat miss-specified, in that it does not fully reflect the true process that generated the nonindependence of the data (Martins and

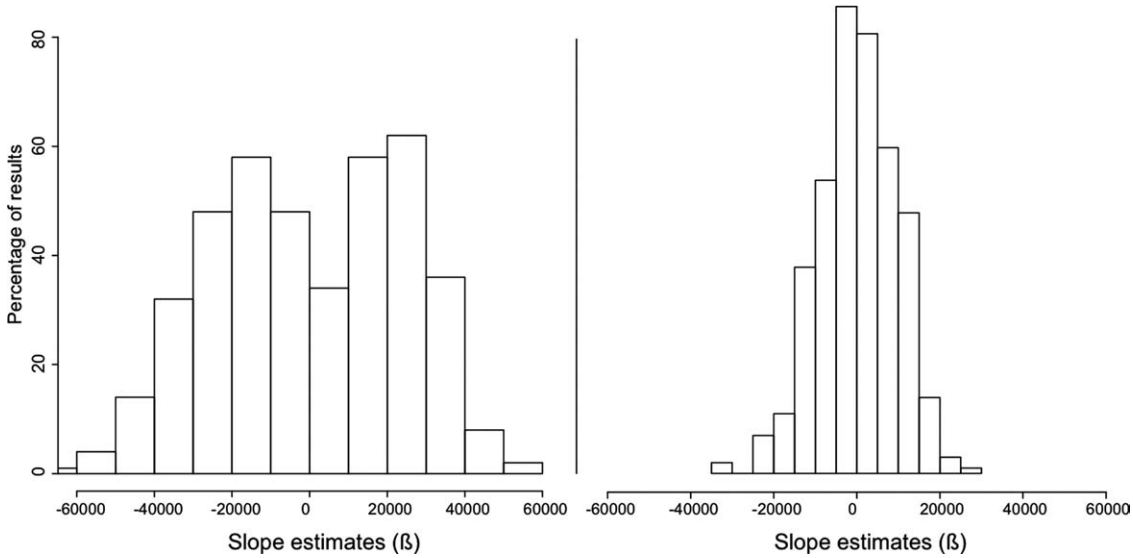


Fig. 3. Histograms of the slope estimates from migratory distance regressed on each of the 405 neutral alleles. The right panel shows the estimates from a least squares model without any adjustment for population data point nonindependence. The left panel shows slopes from the same loci from a model that used our correlation structure to control for the nonindependence that arises from drift and admixture. Note that all estimates are centered on zero (i.e., they are neutral), and that the use of the correlation structure greatly reduces the variance of the estimated slopes.

Garland, 1991; Rohlf, 2006). For example, one factor that may have contributed to miss-specification could be that the D^2 distance does not allow for frequency differences to arise from mutation or from types of selection that may be difficult to detect within the largely neutral markers (e.g., purifying and balancing selection). We expect that future research will enable such miss-specifications to be corrected by using optimizing transformations of the correlation structure based on the dependant and independent variables used in the linear model. Such procedures have recently become routine in the context of phylogenetic generalized least squares (Garland et al., 2005). Some common transformations of the phylogenetic correlation structure to reflect the inferred evolutionary process include the multiplier lambda and the exponential transform rho, both of which control for reduced phylogenetic signal, as well as the exponential transform kappa that reflects speciational evolution (Grafen, 1989; Pagel, 1999). The inflation of Type I error in our study is only slightly greater than observed in phylogenetic simulations before the implementation of such transformations of correlation structures (Martins and Garland, 1991).

Developing appropriate transformations for intraspecific correlation structures is beyond the scope of this study. Thus, we elected to correct the *P*-value from the DRD4 2R/7R result by the inflation of Type I error observed among the neutral loci. The 2R/7R result remained significant after this correction. The residual variation from a least squares regression of migratory distance on DRD4 2 + 7 repeat allele frequency was also significantly less than the residual variation observed among neutral loci.

These results also suggest a strong warning to recent publications by psychologists and others that have heralded the advent of a new "cultural neuroscience" research program (Chiao and Blizinsky, 2009; Chiao et al., 2010; Goldman et al., 2010; Kitayama and Park, 2010; Kitayama and Tompson, 2010; Way and Lieberman, 2010). This program is based on finding associations of gene frequencies with behavioral and ecological traits across human groups but does not incorporate the effects of neutral genetic processes. One example, published recently in AJPA, purported to find evidence of NS alleles of DRD4 being linked to hunter-gatherer lifestyles among Native South Americans (Tovo-Rodrigues et al., 2010). We agree that such a research program, properly conducted, will yield important scientific insights. None of these studies, however, have accounted adequately for the nonindependence of human populations. Tovo-Rodrigues et al. (2010) made no accommodation for the role of neutral evolution of allele frequencies. Both Chiao and Blizinsky (2009) and Way and Lieberman (2010) attempted to deal with nonindependence by lumping data points into similar cultural regions. Anthropologists have demonstrated, however, that such approaches fail to remove the nonindependence that still exists among the lumped cultural regions (Dow, 1984; Mace and Pagel, 1994; Eff, 2004; Dow, 2007). Cultural regions, furthermore, express cultural and linguistic linkages among people and are not designed to reflect the population genetic structure that determines the nonindependence of allele frequencies.

In short, treating human populations as statistically independent treats them as something that profoundly they are not. Populations exhibit complex histories of separation, drift, and admixture, and ignoring this is a serious methodological error that will inflate Type I error, reduce statistical power, and thereby produce many erroneous conclusions.

We found no support for an association between migratory distance and two promoter region loci at DRD4 that alter the expression levels of this gene and mediate NS as measured through psychological surveys. Although the inclusion of more data points may change this result, these negative results are not likely the product of low power, as the power to detect the observed simple Pearson correlation of the 2R/7R frequencies and migratory distance was 0.96 (R package "pwr", Champely, 2009). These promoter loci may not, in fact, correlate with dispersal distance if their phenotypic effects are more greatly influenced by gene X gene and gene X environ-

ment interactions than are the effects of the 2R/7R allele located within exon 3. Given that convergent phenotypic change can often reflect nonconvergent underlying genetic changes (Tishkoff et al., 2006), studies like ours may be able to detect only those loci with highly robust phenotypic effects. Though few, such loci are important to medical intervention, because doctors in their clinical practice are able at present to test for genetic variation at target alleles, whereas diagnoses based on complete individual genotyping may still lie far in the future.

We hope our positive results for DRD4 exon 3 will motivate more research on this topic that will include sampling of both functional and neutral markers from human populations of known provenience. Unfortunately, no such sample exists at present for DRD4 alleles nor for other loci that would be good candidates for selection by migration effects, specifically the serotonin transporter 5-HTT and monoamine oxidase A (Nakamura et al., 2000; Lea and Chambers, 2007). The paucity of data has health consequences because all these loci are linked to a variety of psychopathologies (Caspi et al., 2003; Canli and Lesch, 2007; Guo et al., 2008). Population-level sampling and the resultant better understanding of the functionality of these loci would likely result in better treatment outcomes for patients than do the current common practices of genetic profiling based on broad ethnic labels. Such labels confound genetic and cultural forms of variation (e.g., Munafò et al., 2003; Chiao and Blizinsky, 2009). Genetic and cultural identities frequently overlap, but just as frequently they are disjoint or intersect in complex ways (Risch et al., 2002; Hunley et al., 2009; Long et al., 2009). We can only understand these interactions if we first sample both genetic and cultural systems in ways that are most appropriate to their particular modes of inheritance and distributions of variation.

In summary, we found significant support for dispersal distance during the out-of-Africa migrations being associated with higher frequencies of NS alleles in DRD4 exon 3. Given the findings of Chen et al., (1999) that these alleles did not induce migration within a single generation, the likely best explanation of our observed association is that NS individuals experience higher fitness during migrations. Migration distance, thus, effectively selects for NS phenotypes and associated alleles even though NS individuals may be no more prone to migrate than are other individuals. The atypical pattern of variation in the 7R VNTR of DRD4 exon 3 suggests it emerged as a recent mutational event ~40–50 Ka BP (Ding et al., 2002) and is under positive Darwinian selection (Wang et al., 2004). This coalescence time is roughly coincident with the first observed wave of migration of modern humans from Africa and into Europe and Asia. The coalescence time, genetic signature of selection, and observed association with migratory distance, are all highly consistent with the hypothesis that increased frequencies of 2R and 7R exon 3 VNTRs have been selected by repeated generations of migration.

ACKNOWLEDGMENTS

The authors report no conflicts of interest. The authors thank Patrick McNamara, Charles Nunn, the lab group of Richard Wrangham, and especially Amanda Lobell for helpful discussions on this research. Amanda Lobell and Jason Hodgson provided helpful comments on the manuscript and analysis. The authors also thank Sarah Tishkoff for providing the D^2 distances. The authors thank the Editor-in-Chief, the Associate Editor, and the two anonymous reviewers for their helpful comments on the manuscript.

LITERATURE CITED

Asghari V, Sanyal S, Buchwaldt S, Paterson A, Jovanovic V, Van Tol HH. 1995. Modulation of intercellular cyclic AMP levels by different human dopamine D4 receptor variants. J Neurochem 65:1157–1165.

Canli T, Lesch K-P. 2007. Long story short: the serotonin transporter in emotion regulation and social cognition. Nat Neurosci 10:1103–1109.

Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Braithwaite A, Poulton R. 2003. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science 80:504–517.

Champely, S. 2009. pwr: Basic functions for power analysis. R package version 1.1.1. Available at: http://CRAN.R-project.org/package=pwr.

Chang FM, Kidd JR, Livak KJ, Pakstis AJ, Kidd KK. 1996. The worldwide distribution of allele frequencies at the human dopamine D4 receptor locus. Hum Genet 98:91–101.

Chen C, Burton M, Greenberger E, Dmitrieva J. 1999. Population migration and the variation of dopamine D4 receptor (DRD4) allele frequencies around the globe. Evol Hum Behav 20:309–324.

Chiao JY, Blizinsky KD. 2009. Culture-gene coevolution of individualism-collectivism and the serotonin transporter gene. Proc R Soc B 277:529–537.

Chiao JY, Hariri AR, Harada T, Mano Y, Sadato N, Parrish TB, Iidaka T. 2010. Theory and methods in cultural neuroscience. Soc Cogn Affect Neurosci 5:356–361.

Cloninger CR, Przybeck TR, Svrakic DM. 1991 The tridimensional personality questionnaire: U.S. normative data. Psychol Rep 69(3Part 1):1047–1057.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics 185:1411–1423.

DeGiorgio M, Jakobsson M, Rosenberg NA. 2009. Out of Africa: modern human origins special feature: explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. Proc Natl Acad Sci USA 106:16057–16062.

Deshpande O, Batzoglou S, Feldman MW, Cavalli-Sforza LL. 2009. A serial founder effect model for human settlement out of Africa. Proc Biol Sci 276:291–300.

Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, Schuck S, Swanson JM, Zhang YP, Moyzis RK. 2002. Evidence of positive selection acting at the human dopamine receptor D4 gene locus. Proc Natl Acad Sci USA 99:309–314.

Dow MM. 2007. Galton's problem as multiple network autocorrelation effects–cultural trait transmission and ecological constraint. Cross-Cultural Res 41:336–363.

Dow MM, Burton ML, White DR, Reitz KP. 1984. Galton's problem as network autocorrelation. Am Ethnol 11:754–770.

Dulawa SC, Grandy DK, Low ML, Paulus MO, Geyer MA. 1999. Dopamine D4 receptor-knock-out mice exhibit reduced exploration of novel stimuli. J Neurosci 19:9550–9556.

Ebstein R, Novick O, Umansky R, Priel B, Osher Y, Blaine D, et al. 1996. Dopamine D4 receptor (DRD4) exon III polymorphism associated with the human personality trait of novelty seeking. Nat Genet 12:78–80.

Eff EA. 2004. Does Mr. Galton still have a problem? Autocorrelation in the standard cross-cultural sample. World Cult 15:153–170.

Garland T Jr., Bennet AF, Rezende EL. 2005. Phylogenetic approaches in comparative physiology. J Exp Biol 208:3015–3035.

Garland T Jr., Ives AR. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. Am Nat 155:346–364.

Goldman N, Glei DA, Lin Y-H, Weinstein M. 2010. The serotonin transporter polymorphism (5-HTTLPR): allelic variation and links with depressive symptoms. Depress Anxiety 27:260–269.

Google Inc. Google Earth (Version 5.0.11733.9347).(2009) Mountain View, CA: Google Inc. Available at: http://earth.google.com/.
Grady DL, Chi HC, Ding YC, Smith M, Wang E, Schuck S, Flodman P, Spence MA, Swanson JM, Moyzis RK. 2003. High prevalence of rare dopamine receptor D4 alleles in children diagnosed with attention-deficit hyperactivity disorder. Mol Psych 8:536–545.
Grafen A. 1989. The phylogenetic regression. Phil Trans R Soc Lond B 326:119–157.
Guo G, Ou XM, Roettger M, Shih JC. 2008. The VNTR 2 repeat in MAOA and delinquent behavior in adolescence and young adulthood: associations and MAOA promoter activity. Eur J Hum Genet 16:626–634.
Hancock AM, Witonsky DB, Ehler E, and 8 additional authors. 2010. Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. Proc Natl Acad Sci USA 107:8924–8930.
Harrell FE Jr. 2009. Design: Design Package. R package version 2.3–0. Available at http://CRAN.R-project.org/package=Design.
Hunley KL, Healy ME, Long JC. 2009. The global pattern of gene identity variation reveals a history of long-range migrations, bottlenecks, and local mate exchange: implications for biological race. Am J Phys Anthropol 139:35–46.
Kamakura S, Iwaki A, Matsumoto M, Fukumaki Y. 1997. Cloning and characterization of the 5′ flanking region of the human dopamine D4 receptor gene. Biochem Biophys Res Commun 235:321–326.
Kitayama S, Park J. 2010. Cultural neuroscience of the self: understanding the social grounding of the brain. Soc Cogn Affec Neurosci 5:111–129.
Kitayama S, Tompson S. 2010. Envisioning the future of cultural neuroscience. Asian J Social Psychol 13:92–101.
Klopfstein S, Currat M, Excoffier L. 2006. The fate of mutations surfing on the wave of a range expansion. Mol Biol Evol 23:482–490.
Lea R, Chambers G. 2007. Monoamine oxidase, addiction, and the "warrior" gene hypothesis. N Z Med J 120:1250.
Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319:1100–1104.
Liu H, Prugnolle F, Manica A, Balloux F. 2006. A geographically explicit genetic model of worldwide human-settlement history. Am J Hum Genet 79:230–237.
Long JC, Li J, Healy ME. 2009. Human DNA sequences: more variation and less race. Am J Phys Anthropol 139:23–34.
Lowe N, Kirley A, Mullins C, Fitzgerald M, Gill M, Hawi Z. 2004. Multiple marker analysis at the promoter region of the DRD4 gene and ADHD: evidence of linkage and association with the SNP -616. Am J Med Genet B Neuropsych Genetics 131:33–37.
Mace R, Pagel M. 1994. The comparative method in anthropology. Curr Anthropol 35:349–357.
Martins EP, Garland T Jr. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution 45:534–557.
Mellars P. 2006. Going East: new genetic and archaeological perspectives on the modern human colonization of Eurasia. Science 313:796–800.
Munafò MR, Clark TG, Moore LR, Payne E, Walton R, Flint J. 2003. Genetic polymorphisms and personality in healthy adults: as systematic review and meta-analysis. Mol Psych 8:471–484.
Munafò M, Yalcin B, Willis-Owen S, Flint J. 2008. Association of the dopamine D4 receptor (DRD4) gene and approach-related personality traits: meta-analysis and new data. Biol Psych 63:197–206.
Nakamura M, Ueno S, Sano A, Tanabe H. 2000. The human serotonin transporter gene linked polymorphism (5-HTTLPR) shows ten novel allelic variants. Mol Psych 5:32–38.
Oldenhof J, Vickery R, Anafi M, Oak J, Ray A, Schoots O, Pawson T, von Zastrow M, Van Tol HH. 1998. SH3-binding domains in the dopamine D4 receptor. Biochemistry 37:15726–15736.
Osier MV, Cheung KH, Kidd JR, Pakstis AJ, Miller PL, Kidd KK. 2002. ALFRED: an allele frequency database for anthropology. Am J Phys Anthropol 119:77–83.
Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.
Pinheiro J, Bates D. 2000. Mixed-effects models in S and S-PLUS. New York: Springer Verlag.
Pinheiro J, Bates D, DebRoy S, Sarkar D, andthe R Core team. 2009. nlme: linear and nonlinear mixed effects models. R package version 3.1–96.
Prugnolle F, Manica A, Bolloux F. 2005. Geography predicts neutral genetic diversity of human populations. Curr Biol 15:R159–R160.
Rajeevan H, Osier MV, Cheung KH, Deng H, Druskin L, Heinzen R, Kidd JR, Stein S, Pakstis AJ, Tosches NP, Yeh CC, Miller PL, Kidd KK. 2003. ALFRED–the ALlele FREquency Database–update. Nucleic Acids Res 31:270–271.
Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc Natl Acad Sci USA 102:15942–15947.
Rendall D, Di Fiore A. 1995. The road less traveled: phylogenetic perspectives in primatology. Evol Anthropol 4:43–52.
Reist C, Ozdemir V, Wang E, Hashemzadeh M, Mee S, Moyzis R. 2007. Novelty seeking and the dopamine D4 receptor gene (DRD4) revisited in Asians: haplotype characterization and relevance of the 2-repeat allele. Am J Med Genet Part B Neuropsychiatric Genet 144B:453–457.
Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767–779.
Risch N, Burchard E, Ziv E, Tang H. 2002. Categorization of humans in biomedical research: genes, race and disease. Genome Biol 3: comment 2007.
Rohlf FJ. 2006. A comment on phylogenetic correction. Evolution 60:1509–1515.
Seaman M, Fisher J, Chang F, Kidd KK. 1999. Tandem duplication polymorphism upstream of the dopamine D4 receptor gene (DRD4). Am J Med Genet. Part B Neuropsych Genet 88:705–709.
Sokal RR, Rohlf FJ. 1995. Biometry: the principles and practice of statistics in biological research. New York: W.H. Freeman and Company.
Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbe JB, Awomoyi AA, Bodo JM, Coumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM. 2009. The genetic structure and history of Africans and African Americans. Science 324:1035–1044.
Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghori J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P. 2006. Convergent adaptation of human lactase persistence in Africa and Europe. Nature Genet 39:31–40.
Tomitaka M, Tomitaka S, Otuka Y, Kim K, Matuki H, Sakamoto K, Tanaka A. 1999. Association between novelty seeking and dopamine receptor D4 (DRD4) exon III polymorphism in Japanese subjects. Am J Med Genet 88:469–471.
Tovo-Rodrigues L, Callegari-Jacques SM, Petzl-Erler ML, Tsuneto L, Salzano FM, Hutz MH. 2010. Dopamine receptor D4 allele distribution in Amerindians: a reflection of past behavior differences? Am J Phys Anthropol 143:458–464.
Tsai S, Hong C, Younger W, Chen T. 2004. Association study of catechol-o-methyltransferase gene and dopamine D4 receptor gene polymorphisms and personality traits in healthy young Chinese females. Neuropsychobiology 50:153–156.
Van Tol H, Bunzow JR, Guan HC, Sunahara RK, Seeman P, Niznik HB, Civelli O. 1991. Cloning of the gene for a human dopamine D4 receptor with high affinity for the antipsychotic clozapine. Nature 350:610–614.
Wang E, Ding Y-C, Flodman P, Kidd JR, Kidd KK, Grady DL, Ryder OA, Spence MA, Swanson JM, Moyzis RK. 2004. The genetic architecture of selection at the human dopamine receptor D4 (DRD4) gene locus. Am J Hum Genet 74:931–944.
Way BM, Lieberman MD. 2010. Is there a genetic contribution to cultural differences? Collectivism, individualism and genetic markers of social sensitivity. Soc Cogn Affect Neurosci 5:203–211.

Molecular Psychiatry
https://doi.org/10.1038/s41380-021-01091-4

REVIEW ARTICLE

A unified model of the pathophysiology of bipolar disorder

Paola Magioncalda[1,2,3] · Matteo Martino[1,2]

Received: 7 December 2020 / Revised: 17 March 2021 / Accepted: 29 March 2021
© The Author(s), under exclusive licence to Springer Nature Limited 2021

Abstract

This work provides an overview of the most consistent alterations in bipolar disorder (BD), attempting to unify them in an internally coherent working model of the pathophysiology of BD. Data on immune-inflammatory changes, structural brain abnormalities (in gray and white matter), and functional brain alterations (from neurotransmitter signaling to intrinsic brain activity) in BD were reviewed. Based on the reported data, (1) we hypothesized that the core pathological alteration in BD is a damage of the limbic network that results in alterations of neurotransmitter signaling. Although heterogeneous conditions can lead to such damage, we supposed that the main pathophysiological mechanism is traceable to an immune/inflammatory-mediated alteration of white matter involving the limbic network connections, which destabilizes the neurotransmitter signaling, such as dopamine and serotonin signaling. Then, (2) we suggested that changes in such neurotransmitter signaling (potentially triggered by heterogeneous stressors onto a structurally-damaged limbic network) lead to phasic (and often recurrent) reconfigurations of intrinsic brain activity, from abnormal subcortical–cortical coupling to changes in network activity. We suggested that the resulting dysbalance between networks, such as sensorimotor networks, salience network, and default-mode network, clinically manifest in combined alterations of psychomotricity, affectivity, and thought during the manic and depressive phases of BD. Finally, (3) we supposed that an additional contribution of gray matter alterations and related cognitive deterioration characterize a clinical–biological subgroup of BD. This model may provide a general framework for integrating the current data on BD and suggests novel specific hypotheses, prompting for a better understanding of the pathophysiology of BD.

Introduction

Bipolar disorder (BD) is defined by the occurrence of active episodes of mania and depression, which show opposite constellations of disturbances in psychomotricity, affectivity, and thought, along with asymptomatic periods of euthymia [1, 2]. BD is a prevalent, severe, and debilitating psychiatric disorder, and the understanding of its pathophysiology represents a major challenge for contemporary research in psychiatry and neuroscience [1, 3].

In the last decades, a number of robust and consistent alterations have been detected in BD across diverse areas of research, including immune-inflammatory changes, structural brain abnormalities (in gray and white matter), and functional brain alterations (from neurotransmitter signaling to intrinsic brain activity) [4–9]. However, the relationship between these core alterations in BD is complex and still unclear.

Therefore, in this work, we provide an overview of the most consistent findings on BD, along with the results from our work on this topic, attempting to unify them in an internally coherent working model of the pathophysiology of BD.

Data overview on bipolar disorder

Immunological alterations

Evidences of immune dysregulation and inflammation in psychiatric disorders including BD have been accumulated

These authors contributed equally: Paola Magioncalda, Matteo Martino

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-021-01091-4.

✉ Paola Magioncalda
paola.magioncalda@gmail.com

✉ Matteo Martino
matteomartino9@gmail.com

[1] Graduate Institute of Mind Brain and Consciousness, Taipei Medical University, Taipei, Taiwan

[2] Brain and Consciousness Research Center, Taipei Medical University - Shuang Ho Hospital, New Taipei City, Taiwan

[3] Department of Psychiatry, Taipei Medical University - Shuang Ho Hospital, New Taipei City, Taiwan

Published online: 15 April 2021

since the 1980s, as firstly described by Maes and other Authors, concerning both cytokines and immune cells, in the peripheral circulation and central nervous system [5, 10–15]. In particular, levels or activity of pro-inflammatory cytokines, e.g., interleukin (IL)-6, tumor necrosis factor (TNF)α, interferon (IFN)γ, and other soluble factors such as C-reactive protein (CRP), were found to be consistently elevated in the active phases of BD, mania and depression, with complete or partial normalization during euthymia [16, 17]. Notably, a longitudinal investigation showed that manic episodes were accompanied by state-dependent inflammatory activation (i.e., increased IL-6 and CRP levels) [18]. Moreover, activation of T cells with increased circulating levels of CD4+ T cells and decreased levels of CD8+ T cells, along with activation and increased peripheral levels of monocytes and activation of microglia, have been consistently detected in BD [13, 19–25]. In our previous work, we further specified the T cell alterations in BD, by showing in the peripheral circulation an increase in the early generated CD4+(CD28+) T cells along with a decrease in the differentiated effector CD8+(CD28−) T cells (effector memory, terminal effector memory, and IFNγ-producing CD8+ T cells) [26]. This pattern was strongly associated to BD patients recruited during mania, while less evident in patients recruited during depression or euthymia [26].

Together, these data may suggest that BD is associated to immune activation with a predominant pro-inflammatory profile.

Gray and white matter brain alterations

A large amount of structural neuroimaging data on psychiatric disorders has been accumulated since the 1990s and gray matter (GM) abnormalities were found in BD, although with heterogeneous results [27, 28]. No altered GM volumes were found in first episode and early stage of BD [29, 30]. However, evidence of widespread neuroprogressive loss of GM volume, especially in the frontal areas, has been demonstrated in longitudinal studies and associated with longer illness duration, and this has been related to progressive cognitive deterioration in BD [31, 32].

White matter (WM) abnormalities resulted to be a consistent finding in BD, as shown by several structural imaging studies since the 1990s and diffusion magnetic resonance imaging (dMRI) studies since the 2000s [33]. Increased rate of deep WM hyperintensities was consistently reported in BD [27]. Furthermore, reduced WM volumes were found in BD, even during first episodes [27, 29]. Importantly, widespread WM microstructure alterations (as mainly shown in dMRI data by decreased fractional anisotropy and increased radial diffusivity) were then robustly detected in BD [34–39]. The greater and most consistent WM changes in BD were found in the cingulum (especially anterior cingulum and regions close to the subgenual anterior cingulate cortex, and posterior cingulum) and corpus callosum (especially genu and body), as well as in frontal areas (especially orbitofrontal and subgenual regions), parahippocampal areas, and tracts such as the uncinate fasciculus and fornix, which mostly connect regions of the limbic system [34–39]. In our previous work, we found that widespread WM abnormalities were mainly detectable in patients recruited during the active phases of illness, while WM alterations localized in the anterior midline structures characterized all phases of BD [40]. Moreover, we observed the most consistent WM alterations, as well as deficits of structural connectivity in the anterior midline regions, in patients recruited during mania [26, 41]. Coherently with WM alterations shown in imaging studies, neuropathological data in BD revealed consistent reductions in number or density of oligodendrocytes, along with myelin abnormalities, in anterior brain regions especially [33, 42, 43].

Relationships between white matter and immunological alterations. In our prior work on BD, we firstly detected a specific correlation between WM alterations and the reduction in circulating terminal effector CD8+ T cells (especially in patients recruited during mania) [26]. Notably, such correlation was found only for terminal effector CD8+ T cells, which were reduced in the circulation and are effector cells prone to tissue migration, but not for other cells such as CD4+ T cells [26]. Moreover, WM alterations were also found to correlate with levels of cytokines such as TNFα and IFNγ in BD [44].

Considering all these data, some intriguing similarities as well as specific differences can be noted between BD and multiple sclerosis (MS), the prototype of immune-inflammatory demyelinating diseases of the central nervous system, with potential pathophysiological and psychopathological meanings [5, 26, 45, 46]. Beyond a robust epidemiological association [5, 15, 47–51], a similar immunological pattern can be observed between the two diseases, i.e., the increase in CD4+/CD8+ T cell ratio in the circulation, which is associated with accumulation of effector CD8+ T cells into brain's WM lesions in MS (suggesting that activated CD8+ T cells migrate into the brain where they recognize components of the myelin and contribute to WM damage as effector cells) [19, 26, 52–54]. Coherently, widespread damage in (normal appearing) WM has been reported in both disorders [38, 55, 56]. However, we observed that such WM alterations show a different spatial pattern, with a greater impairment in the anterior brain regions in BD and in the posterior brain regions in MS [57]. Coherently, we also showed that the occurrence of depressive symptomatology in MS was specifically

associated to WM alterations in a cluster mainly encompassing basal ganglia and thalamic areas (including the fornix) that are part of the limbic system, along with a related functional dysconnection of neurotransmitter-related brainstem nuclei [58]. These findings support the possibility of a partial mechanistic overlap in the immune-mediated WM damage between BD and MS, along with the potential role for a specific spatial pattern of WM alterations in the emergence of psychiatric symptomatology.

Collectively, these data may suggest that BD is associated with an immune response sustained by CD4+ T cells that lead to activation of CD8+ T cells and related inflammatory state, and such activated effector CD8+ T cells may migrate into brain tissue where their cytotoxic effects might contribute to widespread WM damage, which most consistently affects tracts connecting regions of the limbic system.

Limbic network alterations

The limbic network (LN) includes the subgenual anterior cingulate cortex and orbitofrontal cortex, along with their connections to the neurotransmitter-related brainstem nuclei, amygdala, hypothalamus, anterior thalamus, and hippocampus, and is involved in the stress-related modulation of homeostasis and neurotransmitter signaling [59–62]. Since the late 1990s, alterations in the LN have been investigated and consistently detected in depressive and bipolar disorders by Drevets, Ongur, Savitz, Mayberg, and other Authors [61, 63, 64]. Specifically, a localized decrement of GM volume was found in the subgenual anterior cingulate cortex and orbitofrontal cortex, which was associated with a reduction in glia without an equivalent loss in neurons, in BD [4, 65, 66]. Increased volume of the amygdala, along with volume reductions in the hippocampus, has been repeatedly detected in BD [4]. Moreover, structural network dysconnectivity (e.g., deficits in clustering and nodal efficiency) was consistently and specifically detected across such limbic regions (including orbitofrontal, cingulate, and hippocampal gyri) in BD [67, 68]. Finally, state-dependent changes in the metabolic activity of the subgenual anterior cingulate and orbitofrontal cortices, along with increased baseline metabolism in the amygdala and hyperactivity of the hypothalamus–pituitary–adrenal (HPA) axis (especially during mania), have been robustly associated with BD [4, 64, 69]. Importantly, the structural damage in the LN in BD seems to be related to metabolic overactivity [4]. Accordingly, inhibition of overactive limbic regions (e.g., inducing a metabolic decrease in the subgenual anterior cingulate cortex and related areas via chronic deep brain stimulation) was found to produce a clinical improvement of depressive symptomatology [64, 70].

Together, all these data might suggest that BD is characterized by an immune-mediated WM damage and its specific localization in the LN may play the core role in the pathophysiology of the disorder.

Stress response, inflammation, and changes in neurotransmitter signaling

Alterations of the LN in BD may have a complex role in the interplay between stress response, inflammation, and neurotransmitter dysregulation [5]. Stress-induced activity in the LN is associated with transient increases in dopamine (DA) and norepinephrine signaling, opioidergic signaling, and HPA axis activity, along with changes in serotonin (5HT) metabolism [71–73]. Moreover, stress response (both stress hormones, such as cortisol or norepinephrine, and sympathetic innervation) modulates immune cell activity, by acutely inducing a Th2 shift in the Th1/Th2 balance, while chronically causing low-grade inflammation characterized by increased pro-inflammatory factors such as IL-6, IFNγ, and TNF [5, 74, 75]. Interestingly, stress system activation can be associated with increased vulnerability to pathogens, especially viral diseases, e.g., a reactivation of latent herpesvirus can occur during stress responses [5]. Accordingly, an association between infection with cytomegalovirus (or also toxoplasma gondii) and depressive or bipolar disorders was reported [5]. Thus, viral reactivation, in presence of immune dysregulation, may contribute to inflammation and brain damage [76]. In turn, pro-inflammatory mediators are able to modulate neurotransmitter signaling [5, 77–79]. Specifically, pro-inflammatory cytokines, such as IFNγ, TNFα, and IL-6, enhance the activity of the indoleamine 2,3-dioxygenase (IDO) enzyme, which diverts tryptophan away from 5HT synthesis and favors the kynurenine pathway [5, 77–79]. This results in decreased availability of 5HT along with production (by activated microglia) of various metabolites, such as the quinolinic acid (able to activate NMDA receptors, potentially inducing excessive glutamate signaling and related excitotoxicity) and kynurenic acid (an NMDA receptor antagonist) [5, 77]. In turn, changes in frontal glutamatergic NMDA-mediated signaling may alter both DA and 5HT signaling (which are mainly reduced or increased by NMDA receptor agonists or antagonists, respectively) [80]. Furthermore, TNF and IL-1 induce the mitogen-activated protein kinases (MAPKs), which increase the expression and function of monoamine transporters, leading to decreased synaptic availability of neurotransmitters such as 5HT and DA [78]. Accordingly, inflammatory stimuli have been shown to induce a decrease in DA release in the basal ganglia, along with reduced activation of the ventral striatum and related emotional blunting, in both healthy and depressed individuals [78].

Coherently, increased levels of inflammatory markers, possibly along with related DA reductions, were associated with functional dysconnection between ventral striatum and ventromedial prefrontal cortex (correlating with anhedonia) and between dorsal striatum and supplementary motor area (correlating with psychomotor inhibition) in depressed patients [81]. However, it is worth noting that decreased availability of 5HT has been shown to induce a functional dysconnection of the 5HT-producing brainstem nuclei only in patients with affective disorders (and not in healthy subjects) [82].

Altogether, these data suggest that stress response and inflammation are physiologically able to induce changes in neurotransmitter availability, and this could pathologically affect the functional architecture of brain activity in BD, potentially in relation to structural alterations of the LN.

Intrinsic brain activity alterations

The functional architecture of intrinsic brain activity revealed various alterations in BD, as shown by resting-state functional MRI studies since the 2000s. Abnormalities in the subcortical–cortical functional connectivity (FC) have been associated with BD, as firstly reported by Anand et al. [83]. Moreover, increased FC between thalamus and sensorimotor cortical areas (along with decreased FC between thalamus and associative frontal cortices) was detected in BD by Anticevic et al. [84] and this finding has been consistently replicated [85–87]. At a network level, a functional dysconnectivity within the default-mode network (DMN) in BD was firstly detected by Ongur et al. in 2010 [88]. Subsequently, functional alterations were also reported in the other main networks, including sensorimotor network (SMN), salience network (SN), and central executive network (CEN), in BD (e.g., [89–93]). However, findings are heterogeneous, at times inconsistent, potentially also reflecting the clinical heterogeneity of BD. Accordingly, our group conducted a series of studies on BD investigating its different phases of illness separately (rather than considering BD overall), aiming to detect distinct brain alterations in mania and depression, which may reflect more closely the underlying pathophysiological features.

Neurotransmitter signaling alterations. In our prior work, we preliminary detected a functional dysconnection of neurotransmitter-related nuclei in BD, with distinct patterns in the different phases of illness [87]. The FC between 5HT-related raphe nuclei (RNi) and basal ganglia-thalamic regions was reduced in mania [87]. Additionally, the FC between DA-related substantia nigra (SNc) and basal ganglia-thalamic regions was reduced in depression (with psychomotor inhibition) [87]. These functional data complemented biochemical and behavioral findings on neurotransmitter signaling in mania and depression. Specifically, decreased 5HT transmission in BD, and in the manic phase especially, as well as decreased DA transmission in the depressive phase, have been reported as the most consistent neurotransmitter finding in BD [6, 94]. Coherently, deficits in 5HT activity have been associated with manic-like symptomatology (e.g., behavioral impulsivity), while deficits in DA activity with depressive-like symptomatology (e.g., motor inhibition) [95–99].

Subcortical–cortical loop alterations. In our previous work, we confirmed the thalamus-SMN hyperconnectivity in our BD sample and extended this finding by showing a distinct pattern of alterations in the active phases of illness [87]. Specifically, the thalamus-SMN FC increased by switching from mainly negative connectivity in healthy subjects to mainly positive connectivity in mania, reflecting an abnormal coupling between thalamus and SMN signals [87]. By contrast, the thalamus-SMN FC increased from mainly negative connectivity in healthy subjects to around-zero connectivity in depression (with inhibited psychomotricity), actually reflecting a decrease in the absolute strength of correlation (i.e., dissociation) between thalamus and SMN signals [87]. Moreover, the thalamus-SMN FC was unchanged in euthymia and, notably, was increased in agitated depression (similarly to mania rather than inhibited depression, suggesting a specific association between thalamus-SMN FC and psychomotor alterations) [87]. Additional subcortical–cortical FC alterations were found in the different BD phases by other Authors: increased FC between basal ganglia and thalamus, and between amygdala and SMN areas (along with decreased FC between amygdala and anterior cingulate cortex) was detected in mania, while increased FC between basal ganglia and SN areas was observed in depression [100–102].

Large-scale network alterations. In our prior work, we found altered topographical distribution in the neuronal variability (i.e., temporal variability of BOLD signal fluctuations, an indirect index of neuronal activity) across the cortex of BD patients, with opposite patterns in the active phases of illness [103]. Specifically, the ratio between neuronal variability in the SMN and DMN (in the ultra-low frequency band slow5 [104]) was increased in mania (i.e., the balance between SMN and DMN was tilted toward the SMN), and decreased in depression (i.e., the SMN/DMN balance was tilted toward the DMN), while no changes in this parameter were found in euthymia [103]. Moreover, we observed an abnormal global signal representation in BD, which was greater in SMN areas in mania while in DMN areas in depression [105]. We also showed reduced intra-network FC within the DMN in mania, while reduced FC within the SMN in depression [41, 90, 106]. Finally, we detected distinct alterations in regional homogeneity and degree of centrality (measures of local and global connectivity, respectively) in the BD phases: regional

homogeneity and degree of centrality (in the ultra-low frequency band slow4 [104]) were decreased in the DMN during mania (deficits of higher frequencies in the DMN were associated with their increase in SMN areas); conversely, regional homogeneity was decreased in the SMN during depression [106]. Notably, all these alterations in intrinsic brain activity coherently correlated with manic and depressive symptomatology in opposite way (interestingly, psychomotor hyperactivity positively correlated with SMN connectivity, while distractibility, which can be considered a deficit in internal thought, negatively correlated with DMN connectivity) [41, 90, 103, 105, 106]. Furthermore, network alterations were reported in the various BD phases by other Authors. Decreased FC within the DMN and increased FC within the dorsal attention network were detected during mania [88, 107]. Mania-inducing brain lesions were found to selectively disrupt functional connections of the associative areas, such as the dorsolateral prefrontal and temporal cortices, or the orbitofrontal cortex [108]. Network hyperconnectivity and greater clustering in the superior frontal gyrus (including the premotor cortex), amygdala, and midbrain (encompassing the DA-related SNc and ventral tegmental area, VTA) have been associated with manic symptomatology [109]. Conversely, decreased amplitude of low-frequency fluctuations in SMN areas (and increased in SN areas) were detected in depression [110]. Increased regional homogeneity was observed in DMN regions in depression [111]. Altered clustering and efficiency in the LN and DMN have been associated with depressive symptomatology [109, 112]. Finally, absence of network alterations was mainly reported during euthymia [113].

Together, these data show that mania might be associated with a functional dysconnection of the 5HT-related RNi, an increase in thalamus-SMN FC toward positive values, and a relative increase in SMN activity along with a decrease in DMN activity (coherently with the negative correlation of RNi-related FC with thalamus-SMN FC and SMN activity, as observed in healthy subjects) [87, 103, 114]. Conversely, depression might be associated with a functional dysconnection of the DA-related SNc, a decrease in thalamus-SMN FC toward around-zero values, and a relative decrease in SMN activity along with an increase in DMN activity (coherently with the positive correlation of SNc-related FC with the absolute value of thalamus-SMN FC and SMN activity, as observed in healthy subjects) [87, 103, 114]. Thus, such findings might suggest that the manic and depressive phases of BD are characterized by distinct functional reconfigurations of intrinsic brain activity, from changes in neurotransmitter signaling to abnormal subcortical–cortical coupling and alterations in network balancing.

Unified model of the pathophysiology of bipolar disorder

Based on the reported data, we propose a unified model of the pathophysiology of BD (Fig. 1).

Damage in the limbic network and alteration of neurotransmitter signaling

We hypothesize that the core pathological alteration in BD is an LN damage resulting in alterations of neurotransmitter signaling. Although heterogeneous conditions can lead to such damage, we suppose that the main pathophysiological mechanism is traceable to an immune-mediated WM damage involving the LN connections, which destabilizes the neurotransmitter signaling.

Immune dysregulation. A dysimmune response with a predominant pro-inflammatory profile occurs in BD [5]. According to the pathophysiological model of prototypical dysimmune diseases, environmental factors (e.g., microbial infections, physical or chemical agents) are supposed to trigger, on a susceptible polygenic background, a predominant and chronic pro-inflammatory activation (e.g., via molecular mimicry or bystander activation of pattern recognition receptors); it has been suggested that similar mechanisms also play a role in BD [5, 115].

White matter alterations. A widespread alteration of WM microstructure is associated with BD [37, 40]. The dysimmune response may lead to migration of immune effector cells, such as activated effector T cells, into brain tissue (along with activation of microglia and increase in pro-inflammatory mediators) and consequent cytotoxic activity on oligodendrocytes and myelin may result in widespread WM damage (similarly to what observed in MS) [5, 26, 54, 56, 57].

Immune-mediated white matter damage in the limbic network and destabilization of neurotransmitter signaling. A concomitant stress-related limbic overactivity is associated with BD [4, 5, 64]. This could play a role in the resulting spatial pattern of the immune-mediated WM damage, as characterized by greater involvement of the LN [37, 40, 41, 57]. Hypothetically, an LN hypermetabolism may divert the cerebral blood flow, along with the pro-inflammatory cells/mediators, onto its hyperactive regions. This may result in a (immune-mediated) structural alteration in such (stress-related) LN, mainly encompassing the subgenual anterior cingulate cortex, orbitofrontal cortex, and their reciprocal connections with the subcortical areas and neurotransmitter-related brainstem nuclei [4, 61, 62]. Thus, such structural alterations in the connections of LN may result in correspondent functional alterations of the neurotransmitter-related nuclei [58], destabilizing the neurotransmitter signaling

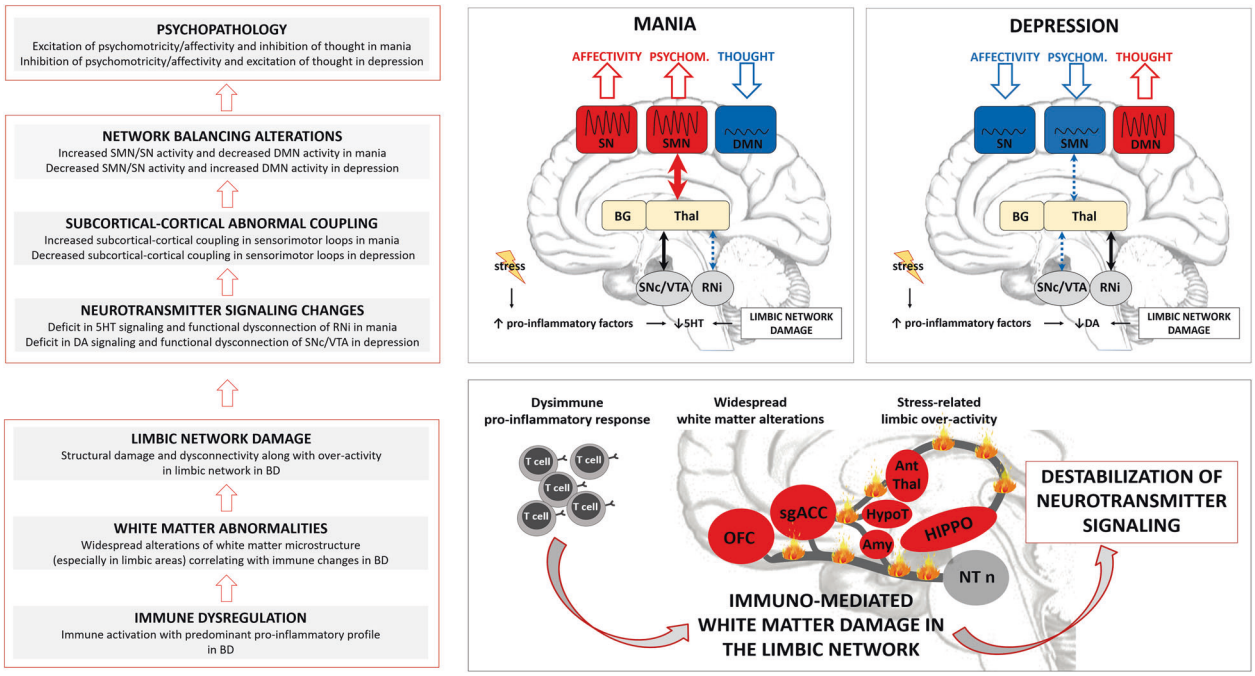


Fig. 1 Unified model of the pathophysiology of bipolar disorder. *Damage in the limbic network and alteration of neurotransmitter signaling (lower part).* According to the model, the core pathophysiological mechanism in BD could be traceable to an immune-mediated white matter alteration resulting in a limbic network damage, which destabilizes the neurotransmitter signaling increasing its susceptibility to perturbations. *Functional reconfiguration of intrinsic brain activity and psychopathology (upper part).* Changes in neurotransmitter signaling (triggered by heterogeneous stressors and inflammatory states onto a damaged limbic network) may then lead to phasic reconfigurations of intrinsic brain activity, from abnormal subcortical–cortical coupling to changes in network balancing. Finally, this may clinically manifest in the psychopathological alterations of mania and depression. *Abbreviations*: BD bipolar disorder, OFC orbitofrontal cortex, sgACC subgenual anterior cingulate cortex, Amy amygdala, HypoT hypothalamus, Ant Thal anterior nuclei of thalamus, HIPPO hippocampus, NT n neurotransmitter-related nuclei, DA dopamine, 5HT serotonin, SNc/VTA substantia nigra/ventral tegmental area, RNi raphe nuclei, BG basal ganglia, Thal thalamus, SN salience network, SMN sensorimotor network, DMN default-mode network.

and increasing its susceptibility to perturbations by various stressors from the external or internal environment.

Functional reconfiguration of intrinsic brain activity and psychopathology

In accordance with our recently proposed three-dimensional model of brain functioning and behavioral/phenomenological patterns [9], we then suggest that changes in neurotransmitter signaling (potentially triggered by heterogeneous environmental stressors onto a structurally-damaged LN) lead to phasic (and often recurrent) reconfigurations of intrinsic brain activity. This might represent a pathological re-setting of the pre-existing baseline in intrinsic brain functioning (and related temperamental features), resulting in distinct changes in network balancing that alter the basal pattern of input/output processing. Finally, this may clinically manifest in the complex psychopathology of manic and depressive states of BD. See Supplementary Material and our previous work [9].

Stress response, inflammation, and changes in neurotransmitter signaling. The occurrence of a stress response, which is facilitated by a hyper-(re)active stress system and able to trigger the active episodes of BD, can cause a prolonged increase in pro-inflammatory factors, such as IL-6, IFNγ, and TNF (e.g., a stress-related Th2 shift in the Th1/Th2 balance may favor a reactivation of latent infection of viruses, like herpesviruses, with related pro-inflammatory rebound) [5, 74]. This, in turn, can induce a reduction of 5HT availability (e.g., via IDO activity and kynurenine pathway enhancement) and/or DA availability (e.g., via MAPK activity enhancement) [5, 77, 78]. A decrease of neurotransmitter availability in a structurally-damaged LN may then lead to a functional dysconnection of the neurotransmitter-related nuclei, finally resulting in functional reconfigurations of intrinsic brain activity.

Changes in neurotransmitter signaling, functional reconfigurations of intrinsic brain activity, and psychopathological states. A reduction of 5HT availability may lead to RNi dysconnection (e.g., [87]). This may result in an abnormal neuronal coupling in the sensorimotor subcortical–cortical loops (e.g., [87]). In turn, this may lead to increased levels of intrinsic activity within the SMN and SN at the expense of the DMN (and related shift toward slow4) (e.g.,

[41, 88, 90, 103, 105–108]). This may over-tune the intrinsic brain activity onto the current environment and clinically manifest in a manic state with excited psychomotricity and affectivity as well as inhibited thought [9]. Conversely, a reduction of DA availability may lead to SNc/VTA dysconnection (e.g., [87]). This may result in a neuronal decoupling in the sensorimotor subcortical–cortical loops (e.g., [87]). In turn, this may lead to decreased levels of intrinsic activity within the SMN and SN with relative increase in the DMN (and related shift toward slow5) (e.g., [103, 105, 106, 110, 111]). This may de-tune the intrinsic brain activity from the current environment and clinically manifest in a depressive state with inhibited psychomotricity and affectivity as well as excited thought [9]. Furthermore, other combinations of changes in neurotransmitter signaling may also occur, resulting in other specific reconfigurations of intrinsic brain activity and related psychopathological states [9, 116]. See Supplementary Material.

Additional pathophysiological factors and clinical–biological subgroups of illness

Finally, we suppose that additional pathophysiological factors may result in clinical–biological subgroups of BD. According to our model, the typical form of BD is primarily characterized by immune-inflammatory factors and related WM damage mainly encompassing the LN. Consequent phasic reconfigurations of intrinsic brain activity might superimpose onto structurally-preserved GM, manifesting in the psychopathological alterations during the active episodes and remission during euthymia. The extent of LN damage could be related to episode recurrence (with lesser or greater LN damage in the occurrence of one/few or multiple episodes, respectively). This may result in the phasic-recurrent course of illness. On the other hand, we suggest that a subgroup of BD might be characterized by additional neurodevelopmental and/or neurodegenerative factors (e.g., alterations in synaptic plasticity) that result in GM loss [4, 31]. Thus, the reconfigurations of intrinsic brain activity might superimpose onto structurally-damaged GM, so that the psychopathological alterations during the active episodes might be associated with impoverished contents of thinking and behavior that persist even during euthymia. The extent of GM damage could be related to cognitive deterioration (with lesser or greater GM damage in the secondary presentation (after recurrent episodes) or primary presentation (since the first episode) of cognitive deterioration, respectively). This may result in a tendency to a progressive course of illness. Thus, such distinct clinical–biological subgroups of BD with predominant relapsing-remitting or progressive patterns (partially mirroring the subtypes of MS) [54, 117] might be related to a predominant contribution of immune-inflammatory or neurodevelopmental/neurodegenerative pathophysiological factors.

See Supplementary Material for a discussion of various critical issues (e.g., [118–120]) and future directions.

Conclusions

Our unified model might provide a framework for integrating the current data on BD and suggest novel specific hypotheses that can be tested in future investigations, prompting for a better understanding of the pathophysiology of BD. In turn, this could provide novel diagnostic markers and therapeutic targets, opening the door to potential new strategies of therapy.

Compliance with ethical standards

Conflict of interest The author declares no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. A.P.A. Diagnostic and statistical manual for mental disorders. 5th ed. (DSM-5). Washington: American Psychiatric Association; 2013.
2. Kraepelin E. Clinical psychiatry. London: Macmillan; 1902.
3. Savitz JB, Rauch SL, Drevets WC. Clinical application of brain imaging for the diagnosis of mood disorders: the current state of play. Mol Psychiatry. 2013;18:528–39.
4. Savitz J, Drevets WC. Bipolar and major depressive disorder: neuroimaging the developmental-degenerative divide. Neurosci Biobehav Rev. 2009;33:699–771.
5. Mechawar N, Savitz J. Neuropathology of mood disorders: do we see the stigmata of inflammation? Transl Psychiatry. 2016;6:e946.
6. Nikolaus S, Antke C, Muller HW. In vivo imaging of synaptic function in the central nervous system: II. Mental and affective disorders. Behav Brain Res. 2009;204:32–66.
7. Menon V. Large-scale brain networks and psychopathology: a unifying triple network model. Trends Cogn Sci. 2011;15:483–506.
8. Northoff G, Hirjak D, Wolf RC, Magioncalda P, Martino M. All roads lead to the motor cortex: psychomotor mechanisms and their biochemical modulation in psychiatric disorders. Mol Psychiatry. 2021;26:92–102.
9. Martino M, Magioncalda P. Tracing the psychopathology of bipolar disorder to the functional architecture of intrinsic brain activity and its neurotransmitter modulation: a three-dimensional model. Mol Psychiatry. 2021; [Online ahead of print].
10. Irwin M, Smith TL, Gillin JC. Low natural killer cytotoxicity in major depression. Life Sci. 1987;41:2127–33.
11. Maes M, Bosmans E, Suy E, Minner B, Raus J. Impaired lymphocyte stimulation by mitogens in severely depressed patients. A complex interface with HPA-axis hyperfunction, noradrenergic activity and the ageing process. Br J Psychiatry. 1989;155:793–8.
12. Anderson G, Maes M. Bipolar disorder: role of immune-inflammatory cytokines, oxidative and nitrosative stress and tryptophan catabolites. Curr Psychiatry Rep. 2015;17:8.
13. Reus GZ, Fries GR, Stertz L, Badawy M, Passos IC, Barichello T, et al. The role of inflammation and microglial activation in the

pathophysiology of psychiatric disorders. Neuroscience. 2015;300:141–54.
14. Kupka RW, Hillegers MH, Nolen WA, Breunis N, Drexhage HA. Immunological aspects of bipolar disorder. Acta Neuropsychiatr. 2000;12:86–90.
15. Barbosa IG, Machado-Vieira R, Soares JC, Teixeira AL. The immunology of bipolar disorder. Neuroimmunomodulation. 2014;21:117–22.
16. Sayana P, Colpo GD, Simoes LR, Giridharan VV, Teixeira AL, Quevedo J, et al. A systematic review of evidence for the role of inflammatory biomarkers in bipolar patients. J Psychiatr Res. 2017;92:160–82.
17. Fernandes BS, Steiner J, Molendijk ML, Dodd S, Nardin P, Goncalves CA, et al. C-reactive protein concentrations across the mood spectrum in bipolar disorder: a systematic review and meta-analysis. Lancet Psychiatry. 2016;3:1147–56.
18. Tsai SY, Chung KH, Chen PH. Levels of interleukin-6 and high-sensitivity C-reactive protein reflecting mania severity in bipolar disorder. Bipolar Disord. 2017;19:708–9.
19. Barbosa IG, Rocha NP, Assis F, Vieira EL, Soares JC, Bauer ME, et al. Monocyte and lymphocyte activation in bipolar disorder: a new piece in the puzzle of immune dysfunction in mood disorders. Int J Neuropsychopharmacol. 2015;18:pyu021.
20. Breunis MN, Kupka RW, Nolen WA, Suppes T, Denicoff KD, Leverich GS, et al. High numbers of circulating activated T cells and raised levels of serum IL-2 receptor in bipolar disorder. Biol Psychiatry. 2003;53:157–65.
21. Tsai SY, Chen KP, Yang YY, Chen CC, Lee JC, Singh VK, et al. Activation of indices of cell-mediated immunity in bipolar mania. Biol Psychiatry. 1999;45:989–94.
22. do Prado CH, Rizzo LB, Wieck A, Lopes RP, Teixeira AL, Grassi-Oliveira R, et al. Reduced regulatory T cells are associated with higher levels of Th1/TH17 cytokines and activated MAPK in type 1 bipolar disorder. Psychoneuroendocrinology. 2013;38:667–76.
23. Brambilla P, Bellani M, Isola M, Bergami A, Marinelli V, Dusi N, et al. Increased M1/decreased M2 signature and signs of Th1/Th2 shift in chronic patients with bipolar disorder, but not in those with schizophrenia. Transl Psychiatry. 2014;4:e406.
24. Kohler O, Sylvia LG, Bowden CL, Calabrese JR, Thase M, Shelton RC, et al. White blood cell count correlates with mood symptom severity and specific mood symptoms in bipolar disorder. Aust N Z J Psychiatry. 2017;51:355–65.
25. Wu W, Zheng YL, Tian LP, Lai JB, Hu CC, Zhang P, et al. Circulating T lymphocyte subsets, cytokines, and immune checkpoint inhibitors in patients with bipolar II or major depression: a preliminary study. Sci Rep. 2017;7:40530.
26. Magioncalda P, Martino M, Tardito S, Sterlini B, Conio B, Marozzi V, et al. White matter microstructure alterations correlate with terminally differentiated CD8+ effector T cell depletion in the peripheral blood in mania: Combined DTI and immunological investigation in the different phases of bipolar disorder. Brain Behav Immun. 2018;73:192–204.
27. Kempton MJ, Geddes JR, Ettinger U, Williams SC, Grasby PM. Meta-analysis, database, and meta-regression of 98 structural imaging studies in bipolar disorder. Arch Gen Psychiatry. 2008;65:1017–32.
28. Birur B, Kraguljac NV, Shelton RC, Lahti AC. Brain structure, function, and neurochemistry in schizophrenia and bipolar disorder-a systematic review of the magnetic resonance neuroimaging literature. NPJ Schizophr. 2017;3:15.
29. Vita A, De Peri L, Sacchetti E. Gray matter, white matter, brain, and intracranial volumes in first-episode bipolar disorder: a meta-analysis of magnetic resonance imaging studies. Bipolar Disord. 2009;11:807–14.
30. Bora E, Fornito A, Yucel M, Pantelis C. Voxelwise meta-analysis of gray matter abnormalities in bipolar disorder. Biol Psychiatry. 2010;67:1097–105.
31. Lim CS, Baldessarini RJ, Vieta E, Yucel M, Bora E, Sim K. Longitudinal neuroimaging and neuropsychological changes in bipolar disorder patients: review of the evidence. Neurosci Biobehav Rev. 2013;37:418–35.
32. Hibar DP, Westlye LT, Doan NT, Jahanshad N, Cheung JW, Ching CRK, et al. Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. Mol Psychiatry. 2018;23:932–42.
33. Mahon K, Burdick KE, Szeszko PR. A role for white matter abnormalities in the pathophysiology of bipolar disorder. Neurosci Biobehav Rev. 2010;34:533–54.
34. Sexton CE, Mackay CE, Ebmeier KP. A systematic review of diffusion tensor imaging studies in affective disorders. Biol Psychiatry. 2009;66:814–23.
35. Vederine FE, Wessa M, Leboyer M, Houenou J. A meta-analysis of whole-brain diffusion tensor imaging studies in bipolar disorder. Prog Neuropsychopharmacol Biol Psychiatry. 2011;35:1820–6.
36. Nortje G, Stein DJ, Radua J, Mataix-Cols D, Horn N. Systematic review and voxel-based meta-analysis of diffusion tensor imaging studies in bipolar disorder. J Affect Disord. 2013;150:192–200.
37. Wise T, Radua J, Nortje G, Cleare AJ, Young AH, Arnone D. Voxel-based meta-analytical evidence of structural disconnectivity in major depression and bipolar disorder. Biol Psychiatry. 2016;79:293–302.
38. Heng S, Song AW, Sim K. White matter abnormalities in bipolar disorder: insights from diffusion tensor imaging studies. J Neural Transm. 2010;117:639–54.
39. Favre P, Pauling M, Stout J, Hozer F, Sarrazin S, Abe C, et al. Widespread white matter microstructural abnormalities in bipolar disorder: evidence from mega- and meta-analyses across 3033 individuals. Neuropsychopharmacology. 2019;44:2285–93.
40. Magioncalda P, Martino M, Conio B, Piaggio N, Teodorescu R, Escelsior A, et al. Patterns of microstructural white matter abnormalities and their impact on cognitive dysfunction in the various phases of type I bipolar disorder. J Affect Disord. 2016;193:39–50.
41. Martino M, Magioncalda P, Saiote C, Conio B, Escelsior A, Rocchi G, et al. Abnormal functional-structural cingulum connectivity in mania: combined functional magnetic resonance imaging-diffusion tensor imaging investigation in different phases of bipolar disorder. Acta Psychiatr Scand. 2016;134:339–49.
42. Savitz JB, Price JL, Drevets WC. Neuropathological and neuromorphometric abnormalities in bipolar disorder: view from the medial prefrontal cortical network. Neurosci Biobehav Rev. 2014;42:132–47.
43. Bellani M, Boschello F, Delvecchio G, Dusi N, Altamura CA, Ruggeri M, et al. DTI and myelin plasticity in bipolar disorder: integrating neuroimaging and neuropathological findings. Front Psychiatry. 2016;7:21.
44. Benedetti F, Poletti S, Hoogenboezem TA, Mazza E, Ambree O, de Wit H, et al. Inflammatory cytokines influence measures of white matter integrity in bipolar disorder. J Affect Disord. 2016;202:1–9.
45. Savitz J. Musings on mania: A role for T-lymphocytes? Brain Behav Immun. 2018;73:151–2.
46. Pape K, Tamouza R, Leboyer M, Zipp F. Immunoneuropsychiatry—novel perspectives on brain disorders. Nat Rev Neurol. 2019;15:317–28.
47. Schiffer RB, Wineman NM, Weitkamp LR. Association between bipolar affective disorder and multiple sclerosis. Am J Psychiatry. 1986;143:94–5.

48. Perugi G, Quaranta G, Belletti S, Casalini F, Mosti N, Toni C, et al. General medical conditions in 347 bipolar disorder patients: clinical correlates of metabolic and autoimmune-allergic diseases. J Affect Disord. 2015;170:95–103.
49. Rosenblat JD, McIntyre RS. Bipolar disorder and immune dysfunction: epidemiological findings, proposed pathophysiology and clinical implications. Brain Sci. 2017;7:144.
50. Turner AP, Alschuler KN, Hughes AJ, Beier M, Haselkorn JK, Sloan AP, et al. Mental health comorbidity in MS: depression, anxiety, and bipolar disorder. Curr Neurol Neurosci Rep. 2016;16:106.
51. Marrie RA, Reingold S, Cohen J, Stuve O, Trojano M, Sorensen PS, et al. The incidence and prevalence of psychiatric disorders in multiple sclerosis: a systematic review. Mult Scler. 2015;21:305–17.
52. Pender MP. CD8+ T-cell deficiency, epstein-barr virus infection, vitamin d deficiency, and steps to autoimmunity: a unifying hypothesis. Autoimmune Dis. 2012;2012:189096.
53. Melzer N, Meuth SG, Wiendl H. CD8+ T cells and neuronal damage: direct and collateral mechanisms of cytotoxicity and impaired electrical excitability. FASEB J. 2009;23:3659–73.
54. Baecher-Allan C, Kaskow BJ, Weiner HL. Multiple sclerosis: mechanisms and immunotherapy. Neuron. 2018;97:742–68.
55. Roosendaal SD, Geurts JJ, Vrenken H, Hulst HE, Cover KS, Castelijns JA, et al. Regional DTI differences in multiple sclerosis patients. Neuroimage. 2009;44:1397–403.
56. Willing A, Friese MA. CD8-mediated inflammatory central nervous system disorders. Curr Opin Neurol. 2012;25:316–21.
57. Piaggio N, Schiavi S, Martino M, Bommarito G, Inglese M, Magioncalda P. Exploring mania-associated white matter injury by comparison with multiple sclerosis: a diffusion tensor imaging study. Psychiatry Res Neuroimaging. 2018;281:78–84.
58. Martino M, Magioncalda P, El Mendili MM, Droby A, Paduri S, Schiavi S, et al. Depression is associated with disconnection of neurotransmitter-related nuclei in multiple sclerosis. Mult Scler. 2020; [Online ahead of print].
59. Purves D, Augustine GJ, Fitzpatrick D, Katz LC, LaMantia AS, McNamara JO. The limbic system, in Neuroscience, 2nd ed. Sunderland, MA: Sinauer Associates; 2001.
60. Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol. 2011;106:1125–65.
61. Drevets WC, Savitz J, Trimble M. The subgenual anterior cingulate cortex in mood disorders. CNS Spectr. 2008;13:663–81.
62. Ongur D, Ferry AT, Price JL. Architectonic subdivision of the human orbital and medial prefrontal cortex. J Comp Neurol. 2003;460:425–49.
63. Drevets WC, Ongur D, Price JL. Neuroimaging abnormalities in the subgenual prefrontal cortex: implications for the pathophysiology of familial mood disorders. Mol Psychiatry. 1998;3:220–6.
64. Mayberg HS. Limbic-cortical dysregulation: a proposed model of depression. J Neuropsychiatry Clin Neurosci. 1997;9:471–81.
65. Drevets WC, Price JL, Simpson JR Jr., Todd RD, Reich T, Vannier M, et al. Subgenual prefrontal cortex abnormalities in mood disorders. Nature. 1997;386:824–7.
66. Ongur D, Drevets WC, Price JL. Glial reduction in the subgenual prefrontal cortex in mood disorders. Proc Natl Acad Sci USA. 1998;95:13290–5.
67. Leow A, Ajilore O, Zhan L, Arienzo D, GadElkarim J, Zhang A, et al. Impaired inter-hemispheric integration in bipolar disorder revealed with brain network analyses. Biol Psychiatry. 2013;73:183–93.
68. O'Donoghue S, Holleran L, Cannon DM, McDonald C. Anatomical dysconnectivity in bipolar disorder compared with schizophrenia: a selective review of structural network analyses using diffusion MRI. J Affect Disord. 2017;209:217–28.
69. Belvederi Murri M, Prestia D, Mondelli V, Pariante C, Patti S, Olivieri B, et al. The HPA axis in bipolar disorder: Systematic review and meta-analysis. Psychoneuroendocrinology. 2016;63:327–42.
70. Mayberg HS, Lozano AM, Voon V, McNeely HE, Seminowicz D, Hamani C, et al. Deep brain stimulation for treatment-resistant depression. Neuron. 2005;45:651–60.
71. Pani L, Porcella A, Gessa GL. The role of stress in the pathophysiology of the dopaminergic system. Mol Psychiatry. 2000;5:14–21.
72. Lanfumey L, Mongeau R, Cohen-Salmon C, Hamon M. Corticosteroid-serotonin interactions in the neurobiological mechanisms of stress-related disorders. Neurosci Biobehav Rev. 2008;32:1174–84.
73. Bali A, Randhawa PK, Jaggi AS. Stress and opioids: role of opioids in modulating stress-related behavior and effect of stress on morphine conditioned place preference. Neurosci Biobehav Rev. 2015;51:138–50.
74. Glaser R, Kiecolt-Glaser JK. Stress-induced immune dysfunction: implications for health. Nat Rev Immunol. 2005;5:243–51.
75. Elenkov IJ. Neurohormonal-cytokine interactions: implications for inflammation, common human diseases and well-being. Neurochem Int. 2008;52:40–51.
76. Zheng H, Ford BN, Bergamino M, Kuplicki R, Hunt PW, Bodurka J, et al. A hidden menace? Cytomegalovirus infection is associated with reduced cortical gray matter volume in major depressive disorder. Mol Psychiatry. 2020; [Online ahead of print].
77. Savitz J. The kynurenine pathway: a finger in every pie. Mol Psychiatry. 2020;25:131–47.
78. Miller AH, Raison CL. The role of inflammation in depression: from evolutionary imperative to modern treatment target. Nat Rev Immunol. 2016;16:22–34.
79. Myint AM, Kim YK. Network beyond IDO in psychiatric disorders: revisiting neurodegeneration hypothesis. Prog Neuropsychopharmacol Biol Psychiatry. 2014;48:304–13.
80. Carlsson A, Waters N, Holm-Waters S, Tedroff J, Nilsson M, Carlsson ML. Interactions between monoamines, glutamate, and GABA in schizophrenia: new evidence. Annu Rev Pharmacol Toxicol. 2001;41:237–60.
81. Felger JC, Li Z, Haroon E, Woolwine BJ, Jung MY, Hu X, et al. Inflammation is associated with decreased functional connectivity within corticostriatal reward circuitry in depression. Mol Psychiatry. 2016;21:1358–65.
82. Salomon RM, Cowan RL. Oscillatory serotonin function in depression. Synapse. 2013;67:801–20.
83. Anand A, Li Y, Wang Y, Lowe MJ, Dzemidzic M. Resting state corticolimbic connectivity abnormalities in unmedicated bipolar disorder and unipolar depression. Psychiatry Res. 2009;171:189–98.
84. Anticevic A, Cole MW, Repovs G, Murray JD, Brumbaugh MS, Winkler AM, et al. Characterizing thalamo-cortical disturbances in schizophrenia and bipolar illness. Cereb Cortex. 2014;24:3116–30.
85. Skatun KC, Kaufmann T, Brandt CL, Doan NT, Alnaes D, Tonnesen S, et al. Thalamo-cortical functional connectivity in schizophrenia and bipolar disorder. Brain Imaging Behav. 2018;12:640–52.
86. Tu PC, Bai YM, Li CT, Chen MH, Lin WC, Chang WC, et al. Identification of common thalamocortical dysconnectivity in four major psychiatric disorders. Schizophr Bull. 2019;45:1143–51.
87. Martino M, Magioncalda P, Conio B, Capobianco L, Russo D, Adavastro G, et al. Abnormal functional relationship of sensorimotor network with neurotransmitter-related nuclei via

subcortical-cortical loops in manic and depressive phases of bipolar disorder. Schizophr Bull. 2020;46:163–74.
88. Ongur D, Lundy M, Greenhouse I, Shinn AK, Menon V, Cohen BM, et al. Default mode network abnormalities in bipolar disorder and schizophrenia. Psychiatry Res. 2010;183:59–68.
89. Khadka S, Meda SA, Stevens MC, Glahn DC, Calhoun VD, Sweeney JA, et al. Is aberrant functional connectivity a psychosis endophenotype? A resting state functional magnetic resonance imaging study. Biol Psychiatry. 2013;74:458–66.
90. Magioncalda P, Martino M, Conio B, Escelsior A, Piaggio N, Presta A, et al. Functional connectivity and neuronal variability of resting state activity in bipolar disorder—reduction and decoupling in anterior cortical midline structures. Hum Brain Mapp. 2015;36:666–82.
91. Doucet GE, Bassett DS, Yao N, Glahn DC, Frangou S. The role of intrinsic brain functional connectivity in vulnerability and resilience to bipolar disorder. Am J Psychiatry. 2017;174:1214–22.
92. Baker JT, Dillon DG, Patrick LM, Roffman JL, Brady RO Jr., Pizzagalli DA, et al. Functional connectomics of affective and psychotic pathology. Proc Natl Acad Sci USA. 2019;116:9050–9.
93. Vargas C, Lopez-Jaramillo C, Vieta E. A systematic literature review of resting state network-functional MRI in bipolar disorder. J Affect Disord. 2013;150:727–35.
94. Shiah IS, Yatham LN. Serotonin in mania and in the mechanism of action of mood stabilizers: a review of clinical studies. Bipolar Disord. 2000;2:77–92.
95. Mosienko V, Beis D, Pasqualetti M, Waider J, Matthes S, Qadri F, et al. Life without brain serotonin: reevaluation of serotonin function with mice deficient in brain serotonin synthesis. Behav Brain Res. 2015;277:78–88.
96. Dalley JW, Roiser JP. Dopamine, serotonin and impulsivity. Neuroscience. 2012;215:42–58.
97. Winter C, von Rumohr A, Mundt A, Petrus D, Klein J, Lee T, et al. Lesions of dopaminergic neurons in the substantia nigra pars compacta and in the ventral tegmental area enhance depressive-like behavior in rats. Behav Brain Res. 2007;184:133–41.
98. van Enkhuizen J, Geyer MA, Minassian A, Perry W, Henry BL, Young JW. Investigating the underlying mechanisms of aberrant behaviors in bipolar disorder from patients to models: rodent and human studies. Neurosci Biobehav Rev. 2015;58:4–18.
99. Cosgrove VE, Kelsoe JR, Suppes T. Toward a valid animal model of bipolar disorder: how the research domain criteria help bridge the clinical-basic science divide. Biol Psychiatry. 2016;79:62–70.
100. Altinay MI, Hulvershorn LA, Karne H, Beall EB, Anand A. Differential resting-state functional connectivity of striatal subregions in bipolar depression and hypomania. Brain Connect. 2016;6:255–65.
101. Brady RO Jr., Masters GA, Mathew IT, Margolis A, Cohen BM, Ongur D, et al. State dependent cortico-amygdala circuit dysfunction in bipolar disorder. J Affect Disord. 2016;201:79–87.
102. Brady RO Jr., Margolis A, Masters GA, Keshavan M, Ongur D. Bipolar mood state reflected in cortico-amygdala resting state connectivity: A cohort and longitudinal study. J Affect Disord. 2017;217:205–9.
103. Martino M, Magioncalda P, Huang Z, Conio B, Piaggio N, Duncan NW, et al. Contrasting variability patterns in the default mode and sensorimotor networks balance in bipolar depression and mania. Proc Natl Acad Sci USA. 2016;113:4824–9.
104. Zuo XN, Di Martino A, Kelly C, Shehzad ZE, Gee DG, Klein DF, et al. The oscillating brain: complex and reliable. NeuroImage. 2010;49:1432–45.
105. Zhang J, Magioncalda P, Huang Z, Tan Z, Hu X, Hu Z, et al. Altered global signal topography and its different regional localization in motor cortex and hippocampus in mania and depression. Schizophr Bull. 2019;45:902–10.
106. Russo D, Martino M, Magioncalda P, Inglese M, Amore M, Northoff G. Opposing changes in the functional architecture of large-scale networks in bipolar mania and depression. Schizophr Bull. 2020;46:971–80.
107. Brady RO Jr., Tandon N, Masters GA, Margolis A, Cohen BM, Keshavan M, et al. Differential brain network activity across mood states in bipolar disorder. J Affect Disord. 2017;207:367–76.
108. Lee I, Nielsen K, Nawaz U, Hall MH, Ongur D, Keshavan M, et al. Diverse pathophysiological processes converge on network disruption in mania. J Affect Disord. 2019;244:115–23.
109. Spielberg JM, Beall EB, Hulvershorn LA, Altinay M, Karne H, Anand A. Resting state brain network disturbances related to hypomania and depression in medication-free bipolar disorder. Neuropsychopharmacology. 2016;41:3016–24.
110. Liu CH, Li F, Li SF, Wang YJ, Tie CL, Wu HY, et al. Abnormal baseline brain activity in bipolar depression: a resting state functional magnetic resonance imaging study. Psychiatry Res. 2012;203:175–9.
111. Liu CH, Ma X, Li F, Wang YJ, Tie CL, Li SF, et al. Regional homogeneity within the default mode network in bipolar depression: a resting-state functional magnetic resonance imaging study. PloS ONE. 2012;7:e48181.
112. Wang Y, Wang J, Jia Y, Zhong S, Zhong M, Sun Y, et al. Topologically convergent and divergent functional connectivity patterns in unmedicated unipolar depression and bipolar disorder. Transl Psychiatry. 2017;7:e1165.
113. Syan SK, Smith M, Frey BN, Remtulla R, Kapczinski F, Hall GBC, et al. Resting-state functional connectivity in individuals with bipolar disorder during clinical remission: a systematic review. J Psychiatry Neurosci. 2018;43:298–316.
114. Conio B, Martino M, Magioncalda P, Escelsior A, Inglese M, Amore M, et al. Opposite effects of dopamine and serotonin on resting-state networks: review and implications for psychiatric disorders. Mol Psychiatry. 2020;25:82–93.
115. Floreani A, Leung PS, Gershwin ME. Environmental basis of autoimmunity. Clin Rev Allergy Immunol. 2016;50:287–300.
116. Rocchi G, Sterlini B, Tardito S, Inglese M, Corradi A, Filaci G, et al. Opioidergic system and functional architecture of intrinsic brain activity: implications for psychiatric disorders. Neuroscientist. 2020;26:343–58.
117. Horrobin DF, Lieb J. A biochemical basis for the actions of lithium on behaviour and on immunity: relapsing and remitting disorders of inflammation and immunity such as multiple sclerosis or recurrent herpes as manic-depression of the immune system. Med Hypotheses. 1981;7:891–905.
118. Zuo XN, Xu T, Milham MP. Harnessing reliability for neuroscience research. Nat Hum Behav. 2019;3:768–71.
119. Gong ZQ, Gao P, Jiang C, Xing XX, Dong HM, White T, et al. DREAM: a toolbox to decode rhythms of the brain system. Neuroinformatics. 2021; [Online ahead of print].
120. Dong HM, Castellanos FX, Yang N, Zhang Z, Zhou Q, He Y, et al. Charting brain growth in tandem with brain templates at school age. Sci Bull. 2020;65:1924–34.

The Journal of Neuroscience, January 1993, 13(1): 87–103

Plasticity in the Frequency Representation of Primary Auditory Cortex following Discrimination Training in Adult Owl Monkeys

G. H. Recanzone,[1] C. E. Schreiner, and M. M. Merzenich

Coleman Laboratory, Departments of Otolaryngology and Physiology and Keck Center for Integrative Neuroscience, University of California at San Francisco, San Francisco, California 94143-0732

Previous studies have shown that the tonotopic organization of primary auditory cortex is altered subsequent to restricted cochlear lesions (Robertson and Irvine, 1989) and that the topographic reorganization of the primary somatosensory cortex is correlated with changes in the perceptual acuity of the animal (Recanzone et al., 1992a–d). Here we report an increase in the cortical area of representation of a restricted frequency range in primary auditory cortex of adult owl monkeys that is correlated with the animal's performance at a frequency discrimination task.

Monkeys trained for several weeks to discriminate small differences in the frequency of sequentially presented tonal stimuli revealed a progressive improvement in performance with training. At the end of the training period, the tonotopic organization of AI was defined by recording multiple-unit responses at 70–258 cortical locations. These responses were compared to those derived from three normal monkeys and from two monkeys that received the same auditory stimuli but that were engaged in a tactile discrimination task. The cortical representation, the sharpness of tuning, and the latency of the response were greater for the behaviorally relevant frequencies of trained monkeys when compared to the same frequencies of control monkeys. The cortical area of representation was the only studied parameter that was correlated with behavioral performance. These results demonstrate that attended natural stimulation can modify the tonotopic organization of AI in the adult primate, and that this alteration is correlated with changes in perceptual acuity.

[Key words: auditory cortex, primate, plasticity, behavior, physiology, discrimination training]

Studies within the somatosensory system have shown that partial peripheral dennervation and restricted natural stimulation can result in a reorganization of the topographic representation of the body surface in the cerebral cortex (see Merzenich et al., 1990; Recanzone and Merzenich, 1992). Similarly, reorganization of the topographic representation of the cochlea in the primary auditory cortex has been demonstrated following partial cochlear lesions (Robertson and Irvine, 1989; Irvine et al., 1991). Tuning properties of single cortical neurons have been shown to be altered by conditioning paradigms (Olds et al., 1972; Disterhoft and Stuart, 1976; Kitzes et al., 1978; Ryugo and Weinberger, 1978) and pharmacological manipulations (Ashe et al., 1989; McKenna et al., 1989; Metherate and Weinberger, 1990; see Weinberger et al., 1990, for review). Although the cortical tonotopic map was not explicitly studied, the functional plasticity at the single neuron level suggests that the tonotopic representation of the cochlea in the cortex may be altered as a result of the behavioral training.

Reorganization of cortical representations has been argued to be continuous throughout life and to reflect an individual's ability to acquire new skills and behaviors (Merzenich et al., 1988, 1990). Improvements in performance have been demonstrated for both simple auditory detection and discrimination tasks (Zwislocki et al., 1958; Sinnott et al., 1985; Prosen et al., 1990), and similar improvements in performance on tactually based tasks have recently been demonstrated to parallel changes in the functional organization of somatosensory cortical areas 3a and 3b (Recanzone et al., 1992a–d). The demonstrations that auditory cortical neurons can alter their frequency-specific response properties by either peripheral dennervation or behavioral conditioning suggested to us that the types of cortical reorganization seen in the somatosensory cortex also occur in primary auditory cortex. Specifically, we hypothesized that the functional organization of AI is altered as a consequence of an animal's improvement in auditory discrimination ability. To test this hypothesis, three adult owl monkeys were trained at an auditory frequency discrimination task for several weeks. The tonotopic representation of the primary auditory cortex was defined electrophysiologically after significant improvements in performance were measured. The response properties of these neurons were then compared to those of normal monkeys, as well as to those of monkeys that had received the same auditory stimulation but were engaged in an unrelated tactile task. The electrophysiological data were then analyzed to determine if any simple measure of the cortical representation of the frequencies used in the behavioral paradigm was correlated with the improved behavioral performance.

Materials and Methods

Animals. Ten adult owl monkeys of both sexes were used. Animals were judged to have normal hearing based on evoked potential recordings (ABR). The outer ear canal was determined to be free of obstruction and infection. Monkeys were given free access to water and were food deprived for a 21–22 hr period prior to each psychophysical testing session. Supplemental fruit was given at the end of each testing session to maintain body weight between 85–95% ad lib. The care and treatment

Received Nov. 27, 1991; revised June 23, 1992; accepted July 9, 1992.

We thank G. T. Hradek, M. L. Sutter, and R. E. Beitel for their important contributions to this work, W. M. Jenkins for insightful discussions regarding the behavioral procedures, and Bret E. Peterson for the computer software used in the data analysis. Funding was provided by NIH Grants NS-10414, GM-0774, and ONR N00014-91-J-1317, Hearing Research Inc., and the Coleman Fund.

Reprint requests should be addressed to M. M. Merzenich at the above address. All other correspondence should be addressed to G. H. Recanzone, Laboratory of Sensorimotor Research, Building 10, Room 10C101, National Institutes of Health, Bethesda, MD 20892.

Copyright © 1993 Society for Neuroscience 0270-6474/93/130087-17$05.00/0

of these monkeys were in accordance with the NIH Guide for the Care and Use of Laboratory Animals (revised 1987).

Psychophysics. The psychophysical apparatus and methods have been described in detail (Recanzone et al., 1991). Five monkeys were trained to detect a difference in the frequency of sequentially presented pairs of tone pips. Monkeys were trained in an acoustically transparent test cage housed within a single-walled acoustic chamber lined with echo-attenuating foam. Auditory stimuli were delivered in the free-field from a single speaker (Realistic model 14-1996) located directly over the animal's head. All aspects of the stimulus generation and data collection were automated and under the control of a microcomputer.

Monkeys were trained to initiate a trial by contacting a hand mold (monkeys OM1, OM2, OM3, CM1) or bar located outside of the test cage (OM4 and OM5), or by pulling a lever located inside the test cage (CM2). In each case, this observing response located the animal's head in a stereotyped location and orientation. Head position was monitored continuously with a video camera. Tone pip pairs were presented every 650 msec. Each tone pip had a duration of 150 msec (3 msec rise/fall), with the onset of the second pip occurring 50 msec after the offset of the first. The frequency of the first tone pip was constant for a given frequency discrimination task. Four different frequency ranges were tested: 2.5 kHz task (2511 Hz standard), 3 kHz task (3188 Hz standard), 5 kHz task (4765 Hz standard), and 8 kHz task (7950 Hz standard). Each monkey was extensively trained on at least one of these four tasks, and most were also tested on at least one other task on a few sessions spaced throughout the training period. The comparison, or S1, stimulus consisted of both tone pips of the pair being equal in frequency (the standard). For the target, or S2, stimulus, the frequency of the second tone pip was different from the frequency of the standard. Monkeys were required to maintain contact with the mold, bar, or lever throughout the presentations of the S1 stimuli, and to release contact upon detection of the S2 stimulus within a 450 msec reward window beginning 100 msec after S2 stimulus onset. Failures to release within this reward window were scored as a "miss" and were punished by a 1–5 sec time-out. Releases before this time were scored as a "false-positive" and also resulted in a time-out. Releases within the reward window were scored as a "hit" and the monkey received a 45 mg banana-flavored food pellet (Bio Serve). Six to ten different S2 stimulus frequencies were presented on at least 20 trials during a given session. These frequencies were chosen such that the monkey received a reward on 65–75% of all trials. Each monkey would perform 400–750 trials in a daily session.

The number of S1 stimulus presentations (bins) on a given trial was pseudorandomly determined to be between 2 and 11. The probability that the next stimulus would be an S2 (approximately 0.3) did not vary during the course of a single trial. In the condition that 11 S1 stimuli were presented, the 12th stimulus was always the S2. These trials were not included in the data analysis, as they were predictably the S2 based on time alone. Trials in which the monkey made a response in bins 1 and 2 were similarly omitted as they were always the S1. The performance (P) for each S2 frequency was defined by

$$P = H \cdot S,$$

where H is the hit rate for that frequency (H = # correct responses/# presentations) and S is the safe rate correction factor (S = 1 − false-positive rate). The psychometric function for each session was determined by plotting P as a function of the frequency difference between the S1 and S2 stimuli (delta frequency: ΔF). Threshold was defined from this function as the S2 frequency at which $P = 0.50$. This is a reliable measure for discrimination performance that is essentially unaffected by false-positive rates below 15% (see Recanzone et al., 1991, 1992a). The slope of the psychometric function was defined as the slope of the line connecting the data points immediately above and below threshold. Threshold was defined in all sessions in which at least one S2 stimulus had a performance value below 0.50 and the false-positive rate for the session was below 15%. The psychophysical data for several different S2 frequencies were subjected to signal detection analysis as described by Green and Swets (1966) for five consecutive training sessions spaced throughout the total training period (see Results). This analysis measures the difference in the response between the S1 and S2 stimuli in units of standard deviation and is independent of changes in the subject's internal criteria.

Three classes of control trials vertified that responses were based on the perception of the frequency of the auditory stimulus. On approximately 2% of the trials, the S2 stimulus was not presented until after 11 S1 presentations (described above). On these trials, a release response after the 12th stimulus always resulted in a reward; however, the performance at any given frequency was comparable regardless of the stimulus presentation bin. On 0.5–1.0% of the trials, the S2 stimulus generated by the computer had a delta frequency of zero (S1 = S2). Monkeys under stimulus control never responded in this condition. Finally, monkeys did not respond on trials in which the speaker was disconnected yet all other aspects of the trials were identical.

In order to prevent intensity cues as a function of frequency consequent to changes in the location of standing waves present in the free field, the intensity of each tone pip was varied independently over a 10–12 dB range. The intensity of the stimuli centered in this range was measured at three or four different frequencies at 12 separate locations throughout the sound field for each of the four tasks (Table 1). These locations were chosen to sample the area occupied by the monkey's head during a session and formed a cube of approximately 5 cm/side, with four additional measurements within this cube. There was no statistically significant difference in the intensity measured throughout the sound field between any of the four tested frequencies for each task (two-tailed t test; $p > 0.10$). Furthermore, repeated measures throughout the course of this study showed no statistical differences over time. The spectral content of these signals was measured in the free field to determine the potential contributions of each harmonic of the signal. The highest-intensity harmonics were always 40–50 dB less than the intensity of the fundamental. Finally, the head position of each monkey was viewed continuously with a video camera on each session, and analysis of head position with the behavioral response was made from several videotaped sessions in each monkey. This analysis showed no consistent relationship between the location or small movements of the monkey's head with the presentation of stimuli or with the behavioral response.

In addition to the five monkeys trained at the auditory frequency discrimination task, one monkey (CM1) was trained to detect a change in the frequency of a sinusoidal tactile stimulus (monkey E-2 of Recanzone et al., 1992a–d). The same auditory stimuli as used in the 8 kHz task were presented simultaneous to the presentation of the tactile stimulus, but did not give any indication of reward availability. Monkey CM2 was trained to perform a tactile vibration detection task in the presence of the same auditory stimuli used in the 5 kHz task. These monkeys represent passive-stimulation controls.

Electrophysiology. The primary auditory cortex was defined electrophysiologically in the seven monkeys trained in a psychophysical task and in three normal owl monkeys (N1, N2, and N3) that had not been trained at any task. Anesthesia was induced with a halothane (3%) : nitrous oxide (72%) : oxygen (25%) gaseous mixture. The femoral vein was cannulated and sodium pentobarbital was given intravenously to maintain an areflexic state of anesthesia (28 mg/kg induction, 1–5 mg/hr maintenance). Penicillin G (30,000 U/24 hr) and atropine sulfate (0.1 mg/kg/12 hr) were given intramuscularly. Lactated Ringer's solution with 5% dextrose was continuously infused (2–5 ml/hr). Heart rate, respiration rate, and other vital signs were monitored. Core body temperature was maintained at 37°C with a thermostatically controlled heating blanket. A craniotomy exposed the relevant cortical region, the dura was removed, and a well of silicon oil was constructed to bathe the cerebral cortex continuously. A 40× image of the cortical surface was taken either photographically or with a digitized video image. All electrode penetrations were parallel to each other. The electrode insertion points were reproduced onto the image of the cortical surface with reference to the vasculature. Electrode penetrations away from the sulcus were made only to a single depth corresponding to the middle cortical layers (500–900 μm below the cortical surface).

The stimulus generation and data acquisition methods have been described in detail elsewhere (Schreiner and Mendelson, 1990). Briefly, single tone pips of 50 msec duration (3 msec on/off ramp) were generated by a microprocessor (TMS32010, 16 bit D/A converter at 120 kHz, low pass filtered at 35 kHz) and presented monaurally via a calibrated headphone (STAX 54) connected to a sound delivery tube sealed into the contralateral ear canal. Characteristic frequency (CF) and threshold were defined audiovisually at all recording locations in all monkeys. Frequency–response areas (FRAs) were defined for all locations in monkeys OM1, OM3, and CM1, and in all three normal monkeys. FRAs were reconstructed from the response to a single stimulus at each of 45 frequencies spanning a 2–5 octave range centered at the approximate characteristic frequency (CF) and equidistant on a logarithmic scale. Each frequency was presented at each of 15 levels spanning a 70 dB range in 5 dB steps (675 total stimuli). Q10dB and Q40dB were defined from these frequency response areas as the CF divided by the bandwidth

Table 1. Mean and SD of sound intensities (dB SPL) measured at 12 locations within the sound field of the apparatus

Task	S1 stimulus	Below threshold	Near threshold	Above threshold
2.5 kHz	64.27 ± 2.7	64.44 ± 2.89	64.17 ± 2.93	65.27 ± 2.93
3 kHz	57.65 ± 3.97	59.55 ± 3.01	60.16 ± 2.90	57.15 ± 3.78
5 kHz	54.72 ± 2.93	55.39 ± 3.08	56.51 ± 3.47	53.55 ± 3.88
8 kHz	71.60 ± 4.69	71.81 ± 4.35		73.13 ± 4.71

Measurements were taken at the mid-range intensity used in the behavioral task. In the 2.5 kHz, 3 kHz, 5 kHz, and 8 kHz tasks, the frequencies measured under "S1 stimulus" were 2511 Hz, 3188 Hz, 4765 Hz, and 7950 Hz; "Below threshold" were 2527 Hz, 3275 Hz, 4849 Hz, and 8036 Hz; "Near threshold" were 2534 Hz, 3367 Hz, 4939 Hz; and "Above threshold" were 2613 Hz, 3641 Hz, 5212 Hz, and 8371 Hz, respectively.

10 dB and 40 dB above minimum threshold, respectively, following methods described by Schreiner and Mendelson (1990) and Sutter and Schreiner (1991). The minimum latency was defined as the time from stimulus onset to the earliest consistent response for all 15 intensity levels presented for the three frequencies nearest CF (45 stimuli total).

Electrodes were parylene-insulated tungsten wires with impedances of approximately 1 MΩ at 1 kHz (Microprobe, Inc). Neural signals were amplified, band-pass filtered (1–10 kHz), and displayed on an oscilloscope and audio amplifier. Spike activity was isolated from this signal with a window discriminator (BAK DIS-1) set to accept all waveforms greater than two times the neural noise level with the time window set for the peak of the largest responses. This eliminated evoked potentials and included approximately one to five independent waveforms of neural origin. The number and arrival time of each event relative to stimulus onset were recorded and stored in a computer (DEC 11/73).

At the start of each experiment, the angle of the electrode was set to be approximately parallel with the angle of the sulcus. Our objective was to define tuning curves limited to neurons in the middle cortical layers, which characteristically have the "best" responses as defined by vigorous activity with a short latency and well-defined tuning curves. Three long penetrations were made approximately 300 μm apart and located along a line perpendicular to the sulcus. In these penetrations, the neural response to acoustic stimuli was defined audiovisually in 500 μm steps. By comparing the neural responses at equivalent distances along these three tracks, we were able to verify that the angle of the penetration was roughly parallel to the sulcus. Electrode tracks near the sulcus were continued until an unambiguous border with another cortical field (marked by an abrupt shift in characteristic frequency) was encountered, or until neural activity could no longer be evoked with acoustic stimuli. These penetrations were commonly 3–4.5 mm in length.

At the conclusion of the electrophysiology experiments, electrolytic lesions were made at selected locations by passing 10 μA of direct current for 10–20 sec. Carbon particle marks were made by insertion of a tungsten electrode that had been coated with india ink at four to seven locations surrounding the investigated cortical tissue. The animals were given a lethal injection of barbiturate, and then perfused intracardially with normal saline at 37°C followed by one of two fixatives in phosphate buffer: 10% formalin or 1.5% paraformaldehyde with 1.5% glutaraldehyde. Frozen sections were cut at 100 μm thickness in the frontal plane. The sections were parallel to the electrode penetrations as indicated by the carbon particle marks. Alternate sections were stained with cresyl violet or for endogenous cytochrome oxidase activity (Wong-Riley et al., 1978). The angle of the electrode penetrations and the distances between recording locations were reconstructed with reference to the electrolytic lesions and carbon particle marks. AI maps represent reconstructions of all recording locations that corresponded to the middle cortical layers III and IV, approximately 500–900 μm below the pial surface. These maps are presented in two dimensions by projecting the location of the recording site along the radial direction to a point where it intersected an imaginary line drawn 700 μm below the cortical surface. This technique limits the distortion that results from projecting recording sites to the surface of the crown of the sulcus. Results from recording locations in other cortical layers are not reported here.

Cortical areas of representation were measured by defining borders as equidistant between neighboring recording locations. The approximate error in these area measurements at interpenetration distances used in these experiments is on the order of 10–20% (see Merzenich et al., 1987; Recanzone et al., 1992b). Statistical analysis of all behavioral and electrophysiological data was done using the software STATVIEW 512+. The specific statistical test was an unpaired one-tailed t test unless otherwise indicated. p values ≤ 0.05 were considered to be statistically significant.

Results

Psychophysics

Improvement in performance with training. The auditory frequency discrimination abilities of each of the five trained monkeys improved progressively with training. The improvement in threshold with training is shown for two representation monkeys in Figure 1. This improvement was characterized by a relatively short and steep phase in which large improvements were recorded between sessions, followed by a longer period in which smaller gains were observed. These observations are summarized for each of the five monkeys in Figure 2. The discrimination threshold was averaged over five sessions at different periods roughly equally spaced throughout the several week training period. In almost every case, the largest gains were measured between the initial five sessions and the five sessions taken after approximately one-third of the total number of sessions. The decrease in threshold was generally less throughout the remainder of training, although significant decreases were measured in every monkey ($p < 0.05$). Final mean discrimination thresholds were slightly better for the 2.5 kHz task than the 8 kHz task (see Table 2). By contrast, the final thresholds for the 5 kHz and the 3 kHz tasks in monkeys OM4 and OM5 were much greater, even though these monkeys also showed significant improvements with training.

Previous studies in owl monkeys on tactile frequency discrimination performance have shown that improvements in minimally detectable frequency differences resulted from (1) a leftward shift in the psychometric function and (2) an increase in the slope of the psychometric function (Recanzone et al., 1992a). Representative psychometric functions from one session within each of the four periods of auditory training shown in Figure 2 are reproduced for monkeys OM1 (*A*) and OM3 (*B*) in Figure 3. Consistent with the results from the tactile task, the performance at each S2 frequency increased in a roughly sequential fashion, with the performance at large frequency differences improving before the performance at smaller frequency differences. For example, delta frequencies (ΔF values) above 300 Hz for OM1 had improved to near maximum performance by session 36, whereas the performance at the ΔF of 80 Hz did not begin to improve until after this session. Similar examples could be cited for each monkey.

The slope of the psychometric function also increased throughout the training period. The average slopes of the psychometric function at threshold over five sessions are shown

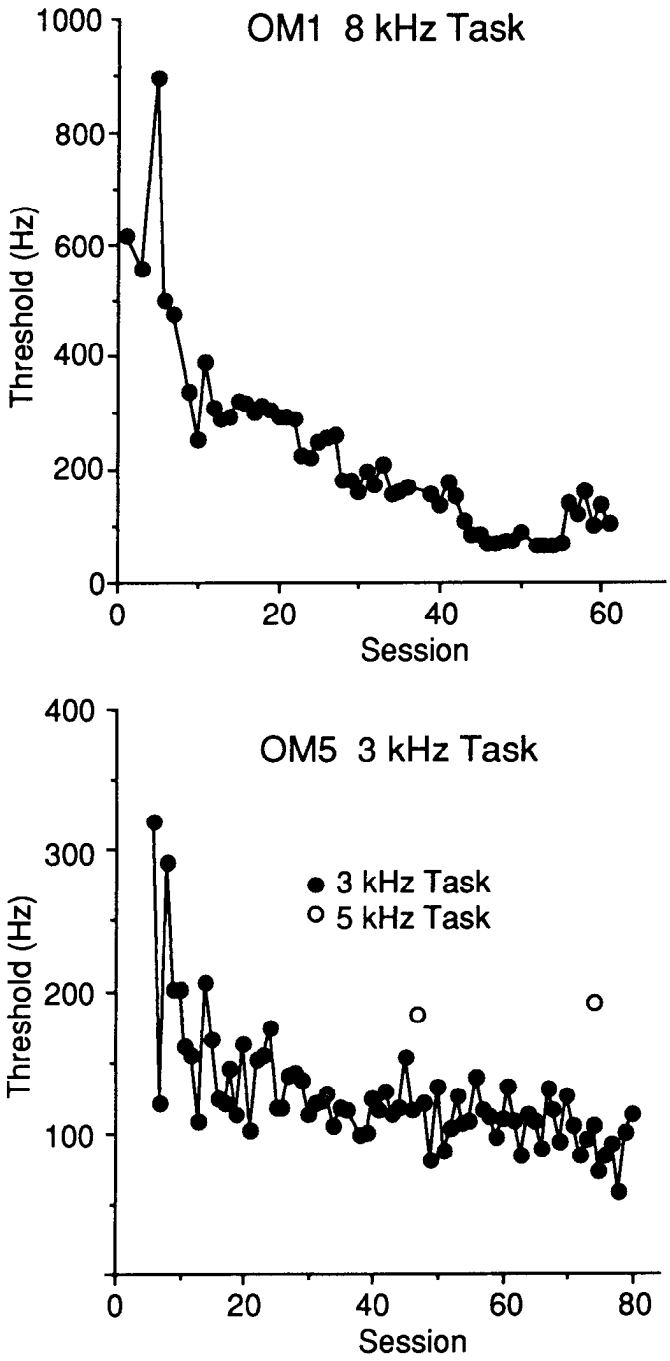


Figure 1. Progressive improvement in the frequency discrimination threshold measured on daily sessions. The delta frequency with a performance of 0.50 is plotted for all sessions in which at least one S2 frequency had a performance below 0.50 (minimum 20 presentations), and the overall false-positive rate was below 15%. *Solid circles* denote the frequency range tested on repeated sessions (8 kHz and 3 kHz for *A* and *B*, respectively), and *open circles* represent sessions using an untrained frequency range (5 kHz in *B*).

for each monkey in Figure 4 for the same training periods as shown in Figure 2. These slopes increased significantly between at least two of the training periods measured, and between the slopes measured in the initial and final training sessions for all monkeys except OM5, which did not reach statistical significance ($p = 0.091$). Thus, consistent with observations in the tactile frequency discrimination task, the improvement in performance was marked by a roughly sequential increase in the

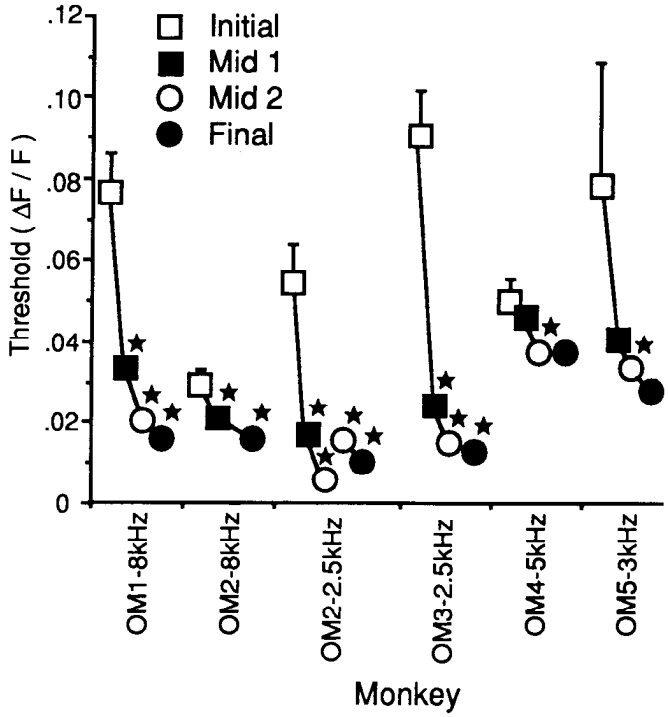


Figure 2. Decrease in frequency discrimination threshold with training. The discrimination thresholds taken over five continuous sessions are shown (mean and SE) for each monkey trained at the auditory frequency discrimination task. Units are given in $\Delta F/F$. *Initial* sessions are taken as sessions 1–5; *Final* sessions are the five sessions immediately preceding the electrophysiology experiment. *Mid 1* and *Mid 2* were chosen to divide the total training period into four roughly equal periods. Error bars not drawn were smaller than the symbol size. *Stars* indicate statistically significant differences between the *starred symbol* and the immediately preceding training period. *OM2-2.5 kHz* is shown in two periods because training at the 2.5 kHz task was interrupted by training on the 8 kHz task (see Fig. 6). Sessions shown for this monkey are 1–5, 30–34, 57–61, 88–92, and 107–111.

performance at each S2 frequency and an increase in the slope of the psychometric function at threshold.

One way in which decreases in threshold could occur is by a shift in the monkey's internal criteria by which it makes a re-

Table 2. Final thresholds measured for all behaviorally trained monkeys

Task	Monkey	# Sessions	$\Delta F/F$
2.5 kHz	OM2	61/24	0.005/0.010
2.5 kHz	OM3	97	0.013
3 kHz	OM5	80	0.028
3 kHz	OM4	6	0.045
5 kHz	OM4	75	0.038
5 kHz	OM5	2	0.040
8 kHz	OM1	61	0.015
8 kHz	OM2	26	0.016
8 kHz	OM3	4	0.041
Passive, 8 kHz	CM1	131	—
Passive, 5 kHz	CM2	75	—

Threshold measurements represent the mean of the final five sessions for trained tasks, or all sessions for tasks with six or less total training sessions. The two measurements for OM2 represent the two training periods for the 2.5 kHz task that bracketed the training at the 8 kHz task. No auditory thresholds were derived in the passive-stimulation animals.

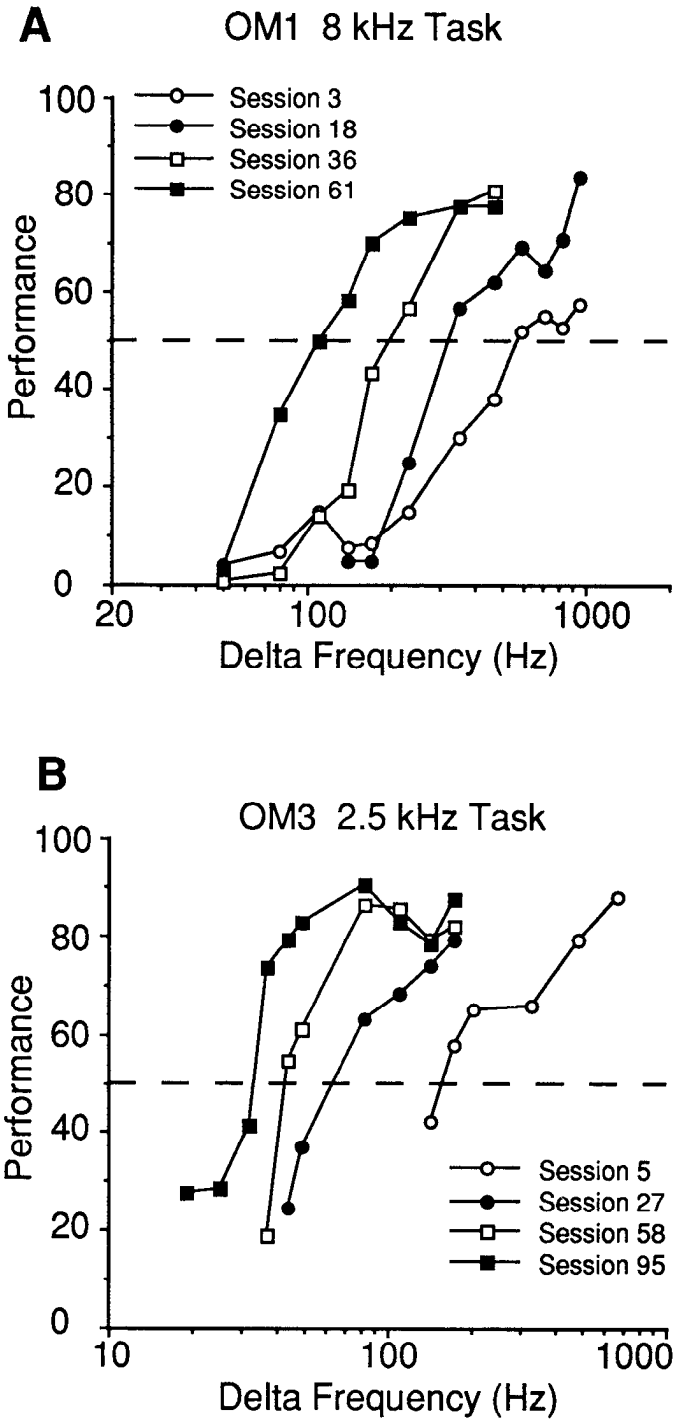


Figure 3. Representative examples of psychometric functions taken during the training periods (see Fig. 2). The *broken line* defines threshold (50% performance). Psychometric functions shifted to smaller frequency differences and appeared to become steeper during the course of training in all studied monkeys.

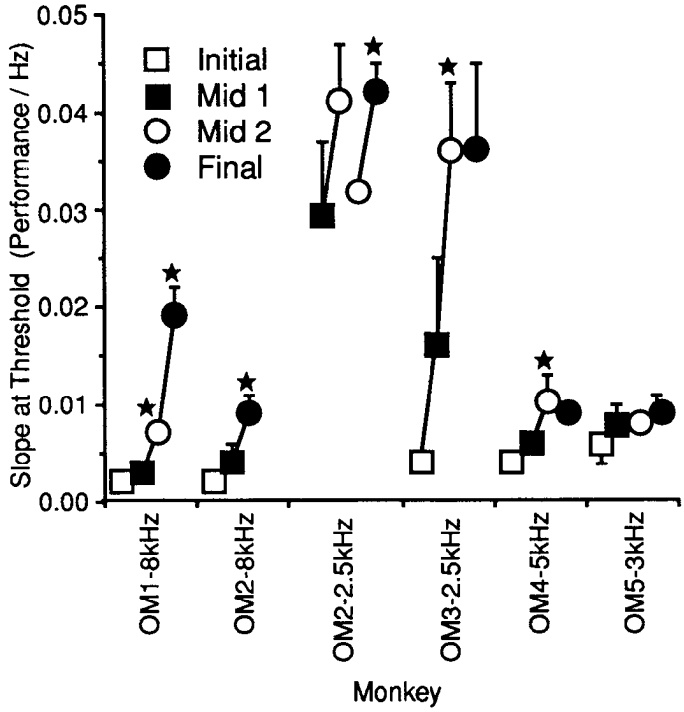


Figure 4. Slope of the psychometric function at threshold in all trained monkeys. The mean and SE of the slope of the psychometric function at threshold are shown for the four training periods as defined in Figure 2 for all monkeys. *Stars* indicate statistically significant differences.

sponse (see Green and Swets, 1966; Recanzone et al., 1992a). The increased slope of the psychometric functions argues against this explanation. A second line of evidence is that the false-positive rate was maintained below 15% and did not significantly change over the course of training. Furthermore, there was no significant correlation between the false-positive rate and the hit rate measured for any ΔF near threshold (hit rates of 20–90%) in any of the five owl monkeys studied ($p > 0.05$ in all cases). Finally, the measure of d' for at least three S2 frequencies was analyzed over the four training periods for each monkey. In every monkey there was at least one statistically significant increase in the measure of d' between training periods for each frequency tested (and no significant decreases). The results from the two monkeys that had the highest discrimination thresholds are shown in Figure 5. Even in these monkeys, there was a sequential increase in the measure of d' with training. These results indicate that the improvements in performance as reflected in the measures of threshold and d' reflect a true increase in perceptual acuity with training.

Transference. Previous studies had indicated that training one sensory surface in a perceptual task confers limited gains in that task to nearby sensory surfaces (see James, 1890; Craig, 1988; Recanzone et al., 1992a). In each monkey excluding OM1, the performance at one other frequency range was tested at various times throughout the training period (Table 2). The most extensively studied animal was OM2, which was trained initially on the 2.5 kHz task, was then shifted to the 8 kHz task, and was finally returned to the 2.5 kHz task (Fig. 6). The characteristic decrease in threshold was observed for both the 2.5 kHz task (solid circles) and the 8 kHz task (open circles). The thresholds measured on the 2.5 kHz task increased during the course of training at the 8 kHz task. The thresholds then decreased again when the task was shifted back to 2.5 kHz. A second feature to note is that the 8 kHz task does not show a rapid improvement in performance at the initial period of training (compare Fig. 6 with Fig. 1, and see Fig. 2).

If a particular frequency outside the normal training range was tested on two or three individual sessions that were widely spaced in time, there was no improvement in performance for

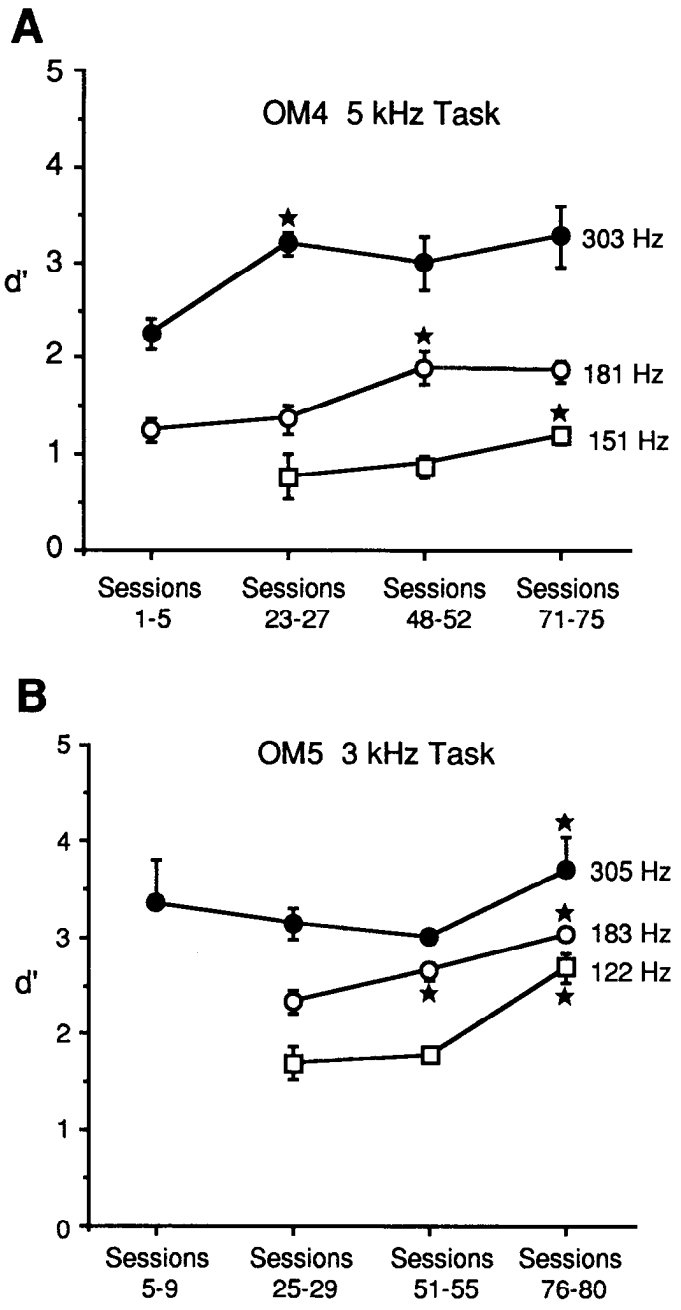


Figure 5. Measures of *d'* for three frequencies as a function of training. The points for a single frequency are connected by a *solid line.* Error bars indicate SE; *stars* indicate statistically significant differences from the preceding measure.

that frequency. This is shown by the performance at 5 kHz in a monkey trained at the 3 kHz task (OM5; open circles, Fig. 1*B*), and by the psychometric functions derived at 3 kHz for a monkey trained at the 5 kHz task (OM4; Fig. 7). In every example, even though there was a significant improvement between these sessions for the trained frequencies, there was little difference in the performance at frequencies that were not trained.

Essentially all training was conducted with the S2 frequencies above the S1 frequency ($+\Delta F$). A somewhat different effect was observed by occasionally testing the performance with decreased S2 frequencies ($-\Delta F$s). On three sessions, OM4 was presented with S2 stimuli in which the second tone of the pair was lower

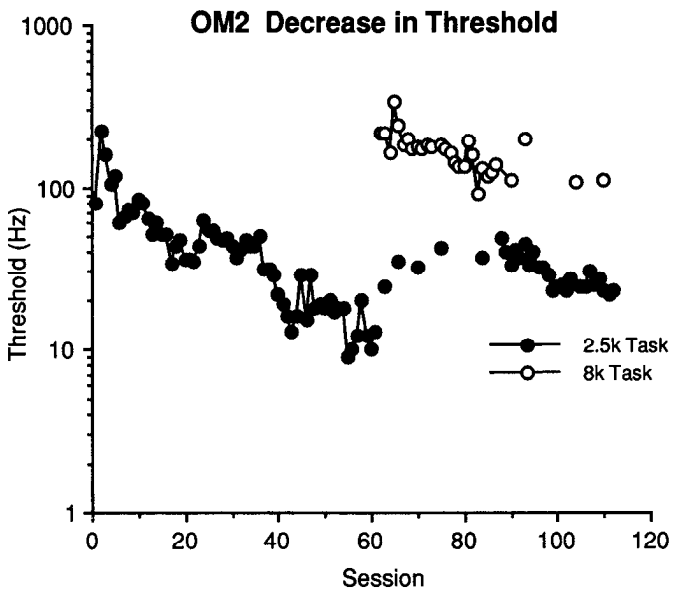


Figure 6. Improvement in frequency discrimination threshold at two different S1 frequencies for monkey OM2. OM2 was initially trained at the 2.5 kHz task (*solid circles*), then changed to the 8 kHz task (*open circles*), and finally returned to the 2.5 kHz task. The 2.5 kHz task was tested intermittently during the consecutive training sessions at the 8 kHz task and vice versa.

in frequency than the first. The psychometric functions from these three sessions are shown in Figure 8 (left), along with three functions using an increased frequency S2 stimulus ($+\Delta F$, right) taken at a similar period of training in the same monkey. The

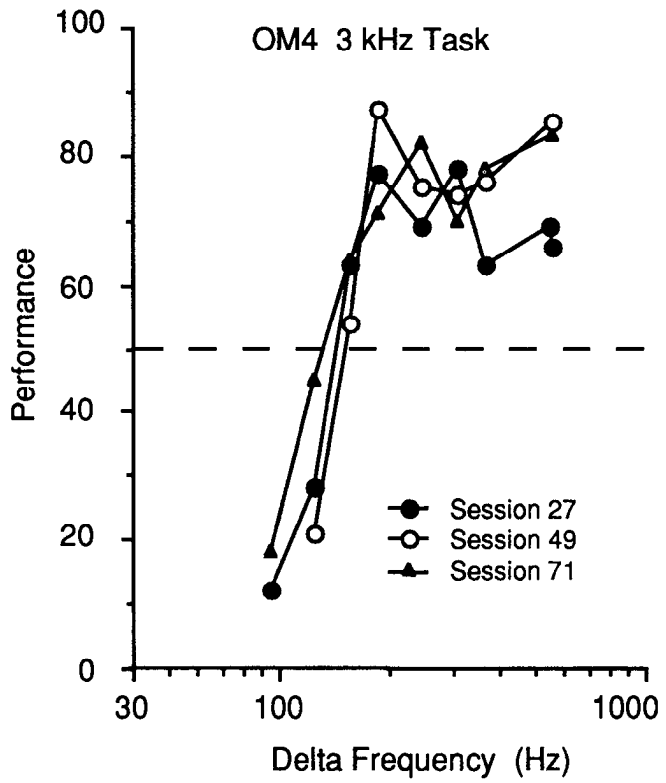


Figure 7. Psychometric functions at three 3 kHz task testing sessions for the monkey trained on the 5 kHz task (OM4). The *broken line* represents performance of 50% (threshold).

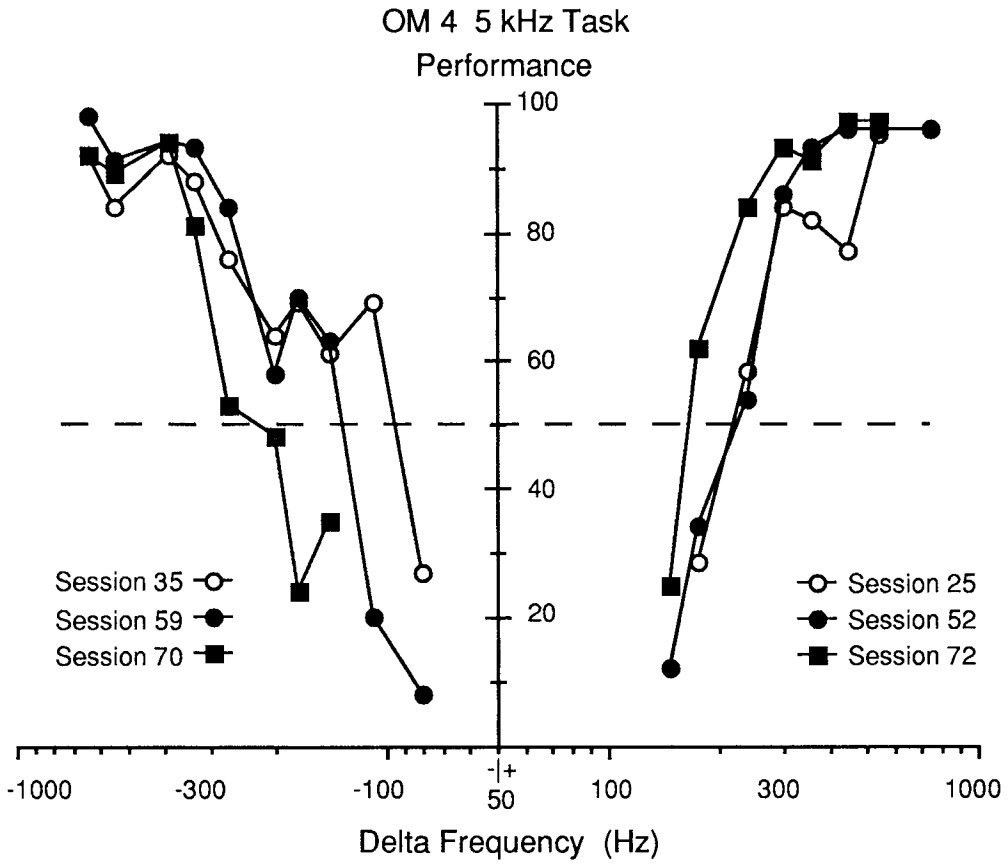


Figure 8. Representative psychometric functions for $+\Delta F$ (*right*) and $-\Delta F$ (*left*) tasks for the 5 kHz standard for monkey OM4. The y-axis intercepts the delta frequency of ±50 Hz. Note the increase in performance for the $+\Delta F$ task and the decrease of performance for the $-\Delta F$ task with training.

functions for the $+\Delta F$ values show the improvement in performance as seen in other monkeys. The functions for $-\Delta F$ values, however, show a worsening of performance and an increase in threshold as a function of session. The worst performance at the $-\Delta F$ task was seen on session 70, in which there were 10 intervening sessions using the increasing delta frequencies. This greatest decrease in performance coincides with an increase in the performance for the $+\Delta F$ stimuli. The same test on the 3 kHz task was done in monkey OM5 near the end of the training period, and the threshold for the $-\Delta F$ task ($\Delta F/F = 0.035$) was greater than that observed for the $+\Delta F$ task ($\Delta F/F = 0.028$).

Effect of stimulus intensity. The effects of transference for other frequencies raise the question of how the performance would be affected for other intensities since the training occurred within a limited intensity range of only 10–12 dB. To address this question, psychometric functions were derived at different intensities in monkeys OM4 and OM5 near the end of the training period. There was a slight decrease in frequency discrimination thresholds with increases in intensity (Fig. 9), but these two parameters were not correlated ($p > 0.15$ in both cases).

Electrophysiology

Functional organization of AI. Basic aspects of the functional organization of AI were determined for both the left and right hemispheres in behaviorally trained monkey OM1 and in one hemisphere of monkeys OM3 and OM4. Sufficient electrophysiological data were not obtained from trained monkeys OM2 and OM5 and will not be discussed further. AI was also defined in the two passively stimulated control monkeys CM1 and CM2, and in three behaviorally naive owl monkeys (N1, N2, and N3). Neural responses in laminae III and IV of AI were marked by short-latency (approximately 10–12 msec) responses to a relatively narrow band of frequencies. In general, the low

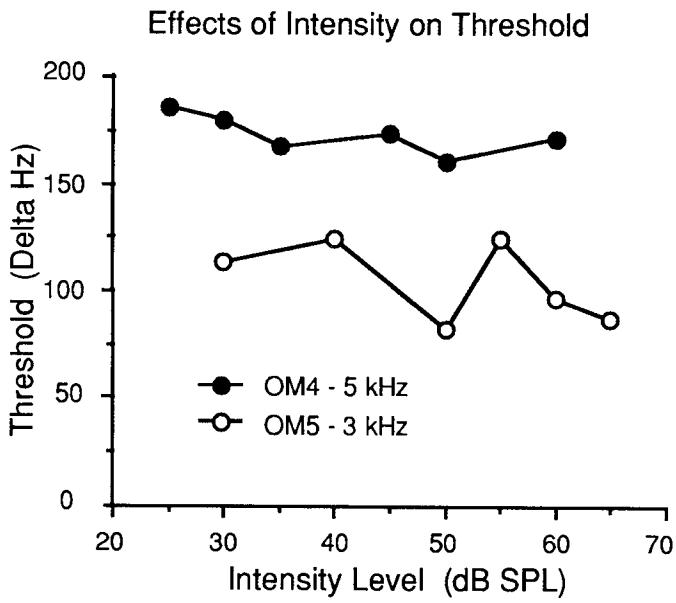


Figure 9. Threshold measured at the trained frequencies as a function of stimulus level for two monkeys, OM4 (*solid circles*) and OM5 (*open circles*). All data were taken from single sessions. The intensity represents the mean intensity measured in the sound field occupied by the monkey's head (see Materials and Methods). Stimulus level varied independently for each tone pip over a range of approximately 15–20% of the mean value plotted.

frequencies were represented on the crown of the gyrus and the higher frequencies were represented down the lateral bank of the sulcus. Neural responses were clustered into "isofrequency bands" where all neurons had similar CFs. These bands were usually oriented oblique to the sulcus but could also be roughly parallel to it. Within the depths of the sulcus, the neural responses would either become very broadly tuned with a higher threshold and longer latency, or become unresponsive to acoustic stimulation. These regions corresponded with the cytoarchitectonic borders of AI as seen histologically, and were characteristic of other auditory cortical fields in this primate species (see Imig et al., 1977; Merzenich et al., 1991).

A central finding of these studies is that the number of recording locations and the cortical area of representation for the frequencies used in the behavioral task were both increased when compared to those frequencies not used in the training. The reconstructed frequency representations from four owl monkeys are shown in Figure 10. Recording locations at which the CF was within the range of frequencies presented in the behavioral tasks are indicated by frequency-specific shading. The total cortical area representing these frequency ranges varied from animal to animal. Some but not all of the differences appeared to be related to the training task. A given frequency was usually represented in a relatively restricted region, but not necessarily continuously. For example, the representation of the frequencies used in the 2.5 kHz task occupied several patches in the trained monkey OM3. By contrast, only a few locations represented these specific frequencies in monkeys not trained at that frequency. Similarly, the greatest representation of the frequencies used in the 5 kHz task were seen in OM4, even though there was only a limited improvement in performance and a relatively small cortical area of representation (Fig. 10*D*). This effect was not seen in the passively stimulated monkey CM2, which received the same acoustic stimulation as OM4, where only a single cortical location represented the frequencies used in the 5 kHz task at CF (Fig. 10*C*). No more than two locations with a CF of 5 kHz were encountered in any of the other monkeys.

The tuning curve data are summarized for all monkeys by plotting the number of cortical locations having a CF within a 500 Hz range (Fig. 11). In Figure 11, the bin size was adjusted to include all the frequencies used in the behavioral tasks (indicated by arrows). In normal owl monkeys, even though all frequencies are represented, there is a consistent underrepresentation of frequencies between 3 and 8 kHz (Fig. 11*A*) and most monkeys showed a relatively large number of recording locations with CFs above 15 kHz. These monkeys also had a large number of locations with a CF near 2 kHz, but those CFs were not in the range used in the 2.5 kHz behavioral task. Each individual monkey showed idiosyncratic increases of particular frequencies, for example, approximately 10 kHz for monkey N2, or approximately 2 kHz for monkey N3.

By contrast, the numbers of locations with CFs in the range of frequencies used in the behavioral task were greatest in monkeys trained at that frequency range. In the trained monkey, the number of locations near 8 kHz was approximately 3–10 times greater in both hemispheres compared to normal and control monkeys (Fig. 11*B*). Monkey CM1, which received the same acoustic stimuli as OM1 but was not required to attend to these stimuli, had an essentially normal representation of these frequencies. The frequency range used in the 2.5 kHz task was represented by more locations in the monkey trained at that task than in any other monkey. The same was true for the frequency range used in the 5 kHz task (Fig. 11*C*).

The larger number of cortical locations that represented the trained frequencies could not be explained by an increase in the sampling density over that region, as the representation of these frequencies and surrounding areas was defined at equivalent density in all monkeys. This increase was better explained by an increase in the cortical area of representation of those frequencies. Figure 12 shows the cortical areas associated with each of the three different frequency ranges for all monkeys. In every case, the area of representation for the behaviorally relevant frequencies for a trained monkey (solid bars) was significantly greater than the cortical areas for those frequencies in all other monkeys, including those that were passively stimulated (hatched bars).

Response properties at individual locations. The tuning properties of the neurons at each recording location were investigated by recording FRAs at all AI recording locations of monkeys OM1, OM3, CM1, and in each of the three normal monkeys. Measures of Q10dB were made at each location; examples are plotted in Figure 13. In every monkey, the values for Q10dB tended to increase with increasing frequency, consistent with reports from other species (see Pelleg-Toiba and Wollberg, 1989; Schreiner and Mendelson, 1990). There was a considerable range of Q10dB values for any given frequency in all monkeys. The Q10dB measures for neurons in trained monkeys were greater (tuning bandwidths were narrower) for the behaviorally trained frequencies when compared to normal or passive stimulation control monkeys. The mean Q10dB and Q40dB for all monkeys over the frequency range used in the 2.5 or 8 kHz tasks are shown in Figure 14*A*. The left and right hemispheres of OM1 are considered independently, and the data from the three normal monkeys are pooled. Measures of Q10dB were statistically significantly greater over the behaviorally trained frequencies in trained monkeys when compared to the three normal controls. The Q10dB of the passive control animals was also elevated over that of normal control monkeys. Statistically significant increases in Q40dB were noted only in the right hemisphere of OM1 for the 8 kHz frequencies.

In addition to the increase in Q10dB, there was also an increase in the minimum latency of the response for the behaviorally trained hemispheres in each of the three cases (Fig. 14*B*). Response latencies in the passive stimulation monkey were not significantly different from those in control monkeys. The variance of the latencies was very low in all monkeys, as indicated by the small SE bars. The distribution of the variance in the latency measure was not statistically significantly different between the normal monkeys and any others ($p > 0.1$).

Correlation of behavioral performance with neural responses. The main objective of these studies was to determine (1) if the functional organization of the primary auditory cortex was altered consequent to training the animal in an acoustic discrimination task, and (2) if the resulting cortical representations were correlated with the behavioral performance. The three measures showing significant changes in trained monkeys described above, the increased cortical area, increased Q10dB (narrower bandwidths near threshold), and increased latency, were subjected to regression analysis. For these analyses, behavioral performance was taken as the threshold measured at the end of the training sessions for the trained monkeys, and paired with the electrophysiological measures derived from those same monkeys. To estimate the behavioral performance of "untrained"

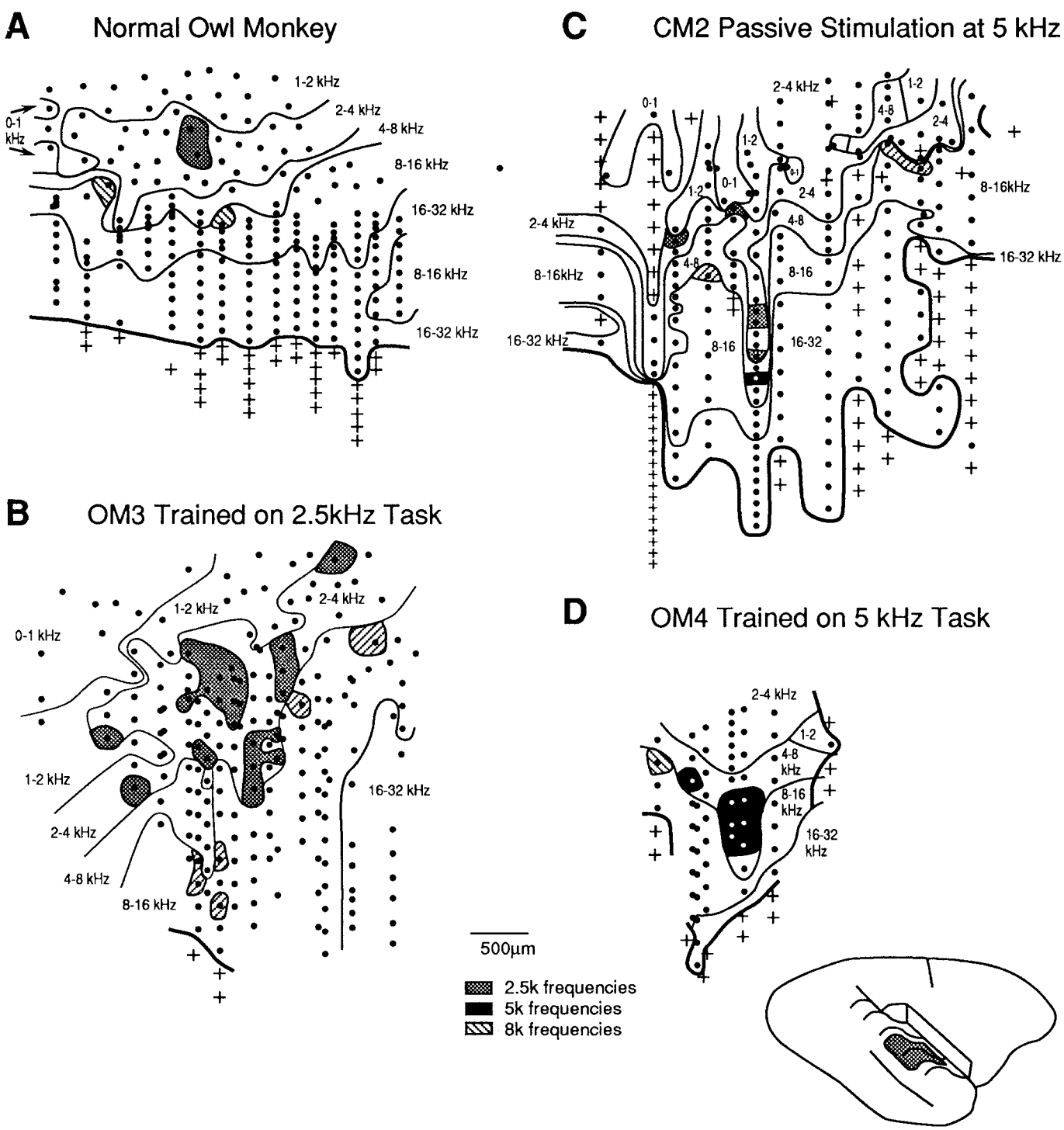


Figure 10. Cortical representations of CF in AI of four adult owl monkeys. *Thin lines* define boundaries of cortical locations with CFs within one octave. *Stippled regions* encircle cortical locations where neurons were recorded with CFs in the frequency range used in the 2.5 kHz task, *solid regions* represent frequencies used in the 5 kHz task, and *hatched regions* represent the frequency range used in the 8 kHz task. The cortical areas representing a given frequency range were approximated by connecting the 50% distance values to the neighboring recording sites with CFs outside the given frequency range. *Pluses* denote recording sites with neuronal responses not consistent with properties of AI neurons. *A* is from a representative normal owl monkey (N2), *B* is from a monkey trained at 2.5 kHz (OM3), *C* shows the monkey passively stimulated with the frequencies used in the 5 kHz task (CM2), and *D* shows the representation of the monkey trained at the 5 kHz task (OM4).

monkeys, the mean threshold of training sessions 5–10 for monkeys OM3 and OM1 were used for the 2.5 kHz and 8 kHz tasks, respectively. These sessions were used because they showed less variability than the initial five sessions (1–5) and probably better reflect the monkey's true initial discrimination abilities (see Discussion). For the "untrained" behavioral performance of the 3 kHz and 5 kHz tasks, the mean thresholds for the two and six sessions at these tasks for monkeys OM4 and OM5 were used, respectively. These "untrained" behavioral measures were considered with the electrophysiological results pooled from the untrained monkeys. The values of Q10dB, Q40dB, and minimum latency showed no significant correlation with these be-

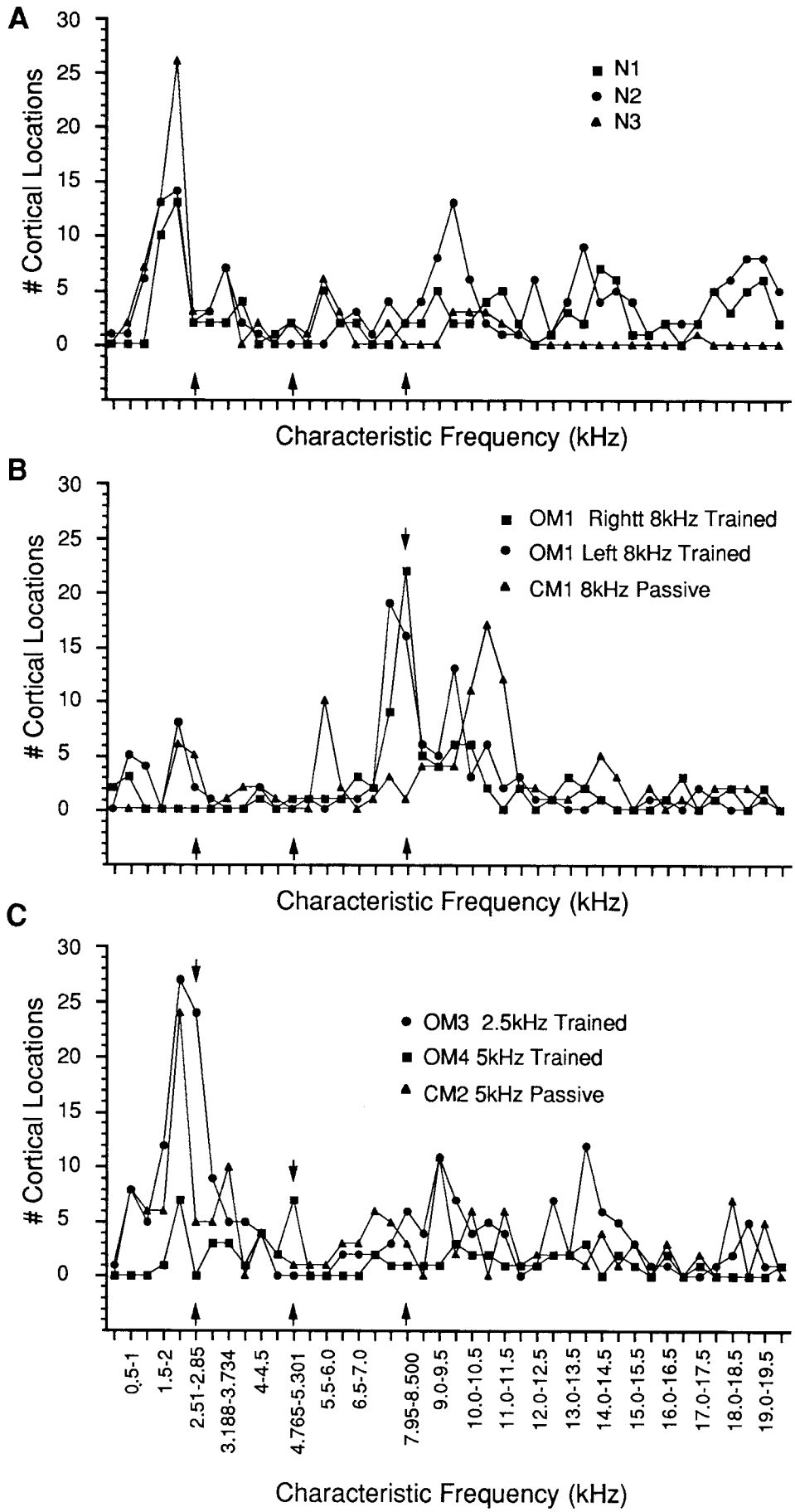


Figure 11. The number of cortical locations plotted by CF in 500 Hz bins. Bin size was adjusted to incorporate all frequencies used in the behavioral task into a single bin (indicated by *arrows*). The three normal monkeys are shown in *A*, the two hemispheres from monkey OM1 and the passive-stimulation control monkey CM1 are shown in *B*, and the monkeys trained at the 5 kHz and 2.5 kHz task (OM4 and OM3) and the passively stimulated monkey CM2 are shown in *C*.

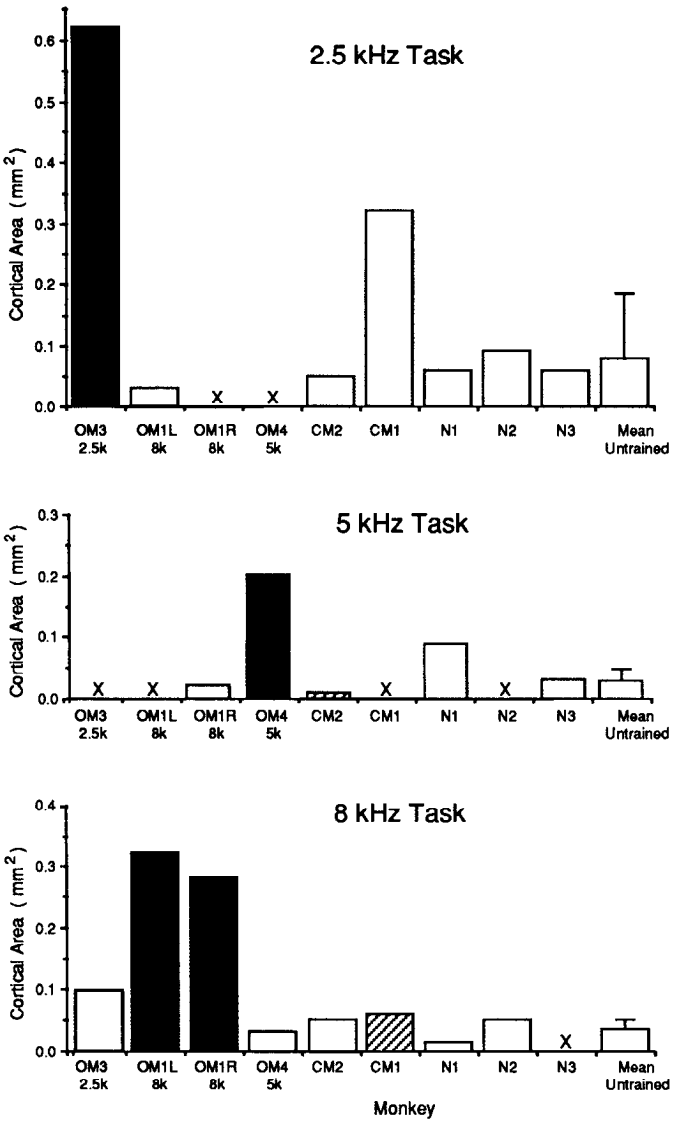


Figure 12. Cortical area of representation of the frequency ranges used in the behavioral study. *Solid bars* indicate the area of representation of the trained frequency for that particular monkey; *hatched bars* represent areas of the frequencies presented to the passive-stimulation control monkeys. *Crosses* represent hemispheres where no cortical locations were recorded with a CF within that frequency range. The column "*Mean Untrained*" represents the mean and SD for all hemispheres excluding the trained hemispheres.

havioral measures (p = 0.44, 0.21, and 0.91, respectively). By contrast, the correlation analysis of behavioral threshold with cortical area revealed that these two parameters were significantly correlated (Fig. 15; r = 0.88; $p < 0.01$; df = 6).

Discussion

Improvements in behavioral performance in an auditory frequency discrimination task were recorded in five adult owl monkeys. There was a roughly sequential improvement in performance at individual S2 frequencies, which was expressed in a shift of the psychometric function and an increase in the slope of the function near threshold. The improvement in performance was much less for untrained frequency ranges, and the performance was independent of stimulus level between 25 and 65 dB SPL. Electrophysiological mapping experiments revealed

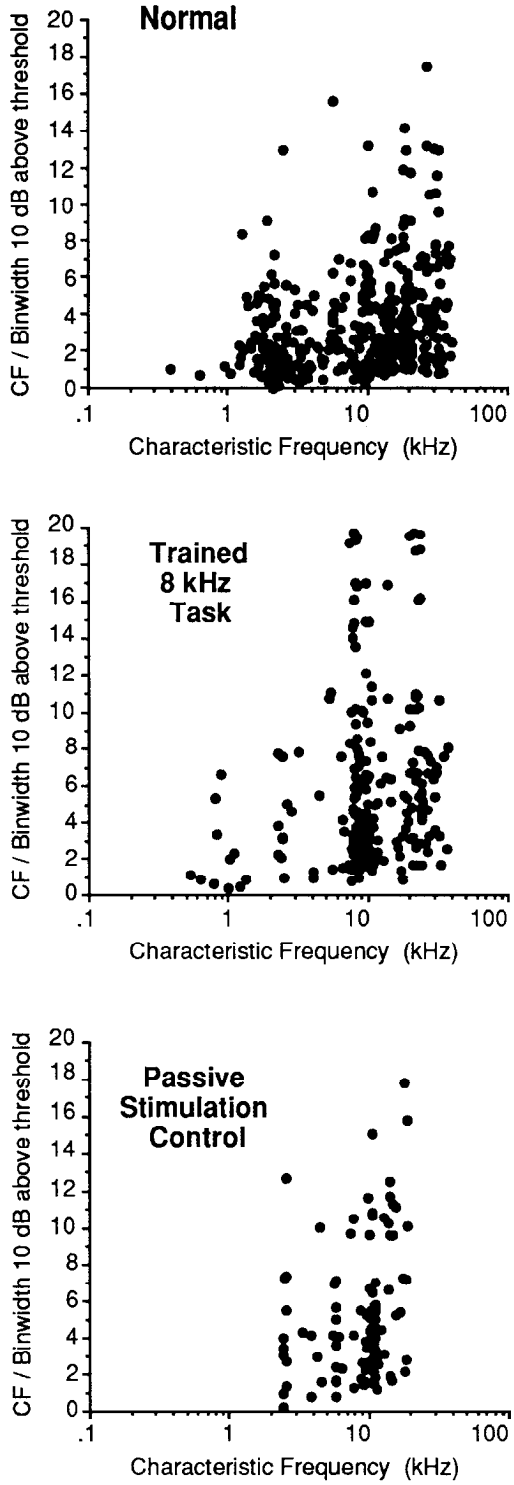


Figure 13. Values of Q10dB as a function of CF. Normal owl monkeys N1, N2, and N3 are pooled and shown in the *top panel*, both hemispheres of trained monkey OM1 in the *middle panel*, and the passive stimulation control monkey CM1 in the *bottom panel*. Each point represents the CF and Q10dB at a single recording location.

a larger cortical area of representation of the behaviorally trained frequencies, sharper tuning as measured by Q10dB but not Q40dB, and longer response latencies over the frequency range used in the behavioral task. The cortical area of CF representation was the only measured parameter that was significantly

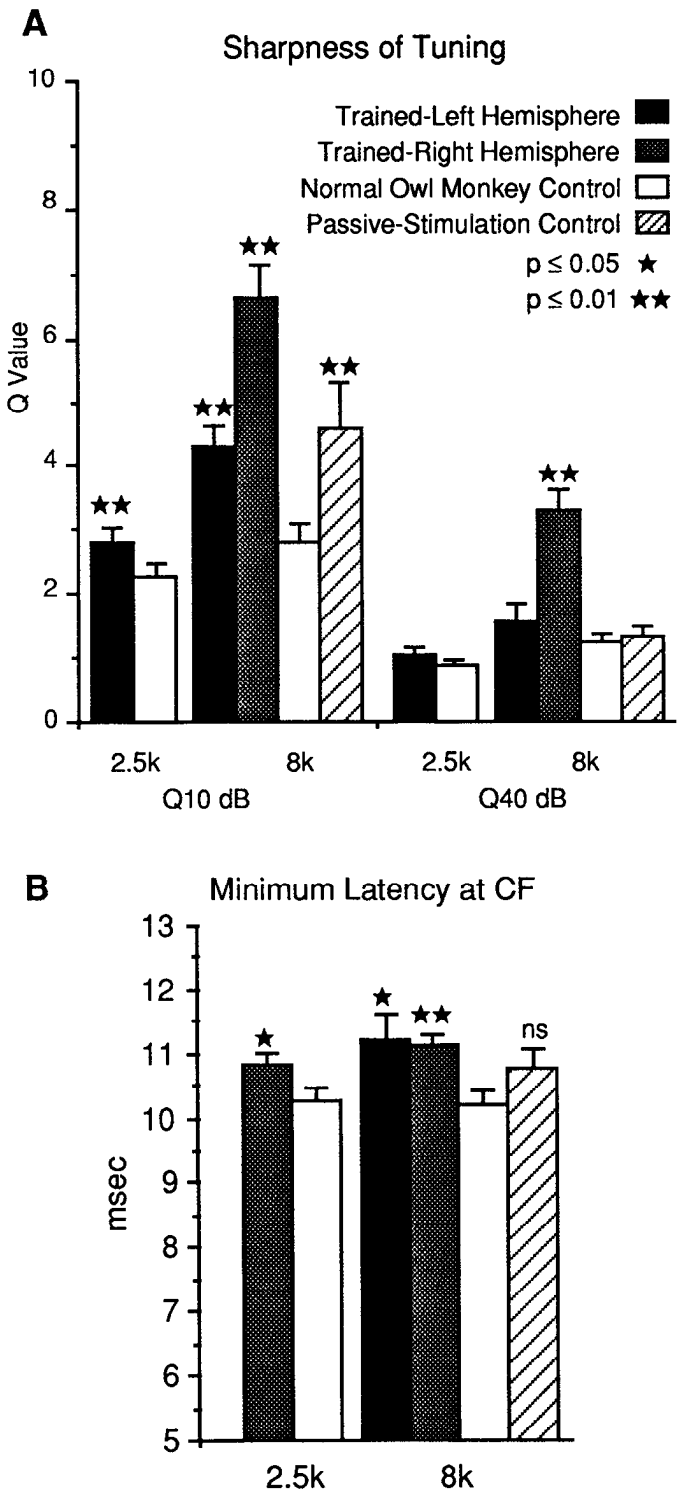


Figure 14. Comparison of sharpness of tuning (*A*) and minimum latency (*B*) for different training conditions. Data from each experimental and passive-stimulation hemisphere are shown as a single bar, with the data from the three normal monkeys combined (*open bars*). *Stars* indicate statistically significant differences from the normal hemispheres. Data are pooled for all recording locations with CFs between 1.9–3.1 kHz (2.5 kHz task) and 6.6–9.9 kHz (8 kHz task). *A*, Mean and SE of Q10dB (*left*) and Q40dB (*right*) for all studied monkeys. *B*, Minimum latency of the response for the 2.5 kHz (*left*) and 8 kHz tasks (*right*).

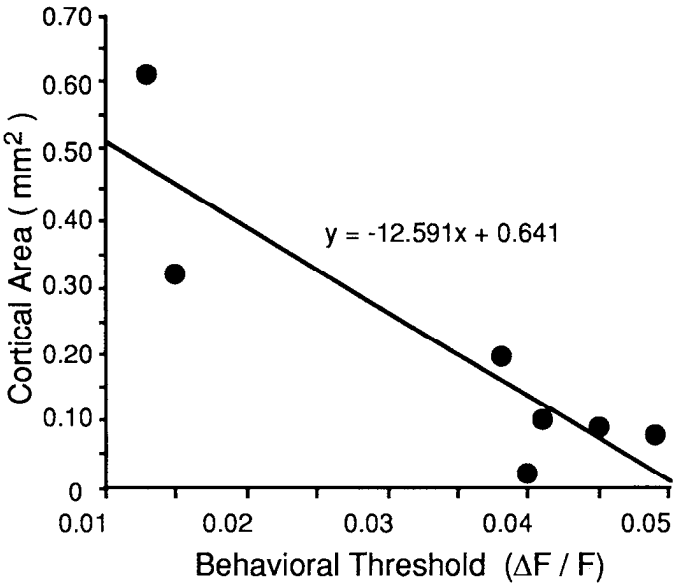


Figure 15. Correlation analysis of the cortical area of representation of the behaviorally trained frequencies at CF with the behavioral threshold. The equation describes the line with the root mean square best fit to the data ($r = 0.882$, $P < 0.01$, df = 6).

correlated with behavioral frequency discrimination performance.

Behavioral thresholds measured in this study were similar to those described for other nonhuman primates for the 2.5 kHz and 8 kHz tasks (Sinnott et al., 1985, 1987; Prosen et al., 1990), but were higher than those recorded in humans (Wier et al., 1977; Sinnott et al., 1985, 1987). The frequency discrimination thresholds measured for the 3 kHz and 5 kHz tasks were greater than those measured for this frequency range in Old-World monkeys. That may be the result of an attenuation of these middle frequencies in this particular species, for example, by physical constraints imposed by the outer or middle ear. These higher-frequency discrimination thresholds are consistent with the relatively high behaviorally measured detection thresholds in the frequency range of 3–6 kHz in this species (Beecher, 1974), the relatively small cortical representation of this frequency range in AI of normal owl monkeys (Imig et al., 1977; present results), and the higher response thresholds of AI neurons over this frequency range in this and other New-World monkey species (Aitkin et al., 1986; Merzenich et al., 1991).

The electrophysiological experiment revealed a tonotopic organization of owl monkey AI similar to that previously shown in this species (Imig et al., 1977) as well as other primates (Merzenich and Brugge, 1972; Aitken et al., 1986). The responses of neuronal clusters to these tonal stimuli were commonly very brisk, with an initial short-latency burst of activity. In spite of the globally smooth tonotopic organization in AI, the representation of very small frequency ranges was often seen to be patchy for both the experimental and control hemispheres. Part of the explanation is that we were basing the area measurements on the CF at threshold, and considered only a very small frequency range. It is likely that this patchy representation would be more continuous if considered from the view of either a higher-intensity stimulus or over a larger frequency range. However, the presence of patches of representation of some frequencies may accurately reflect the tonotopic organization at

this level of detail, as a similarly discontinuous representation of restricted frequency ranges can be seen in the detailed maps of AI described in the marmoset (Aitkin et al., 1986).

Neural mechanisms of frequency discrimination. The finding that the cortical area of CF representation was correlated with the behavioral performance does not restrict the site of frequency discrimination to that location. Most nuclei throughout the central auditory pathways could contribute to these perceptual distinctions. Studies have indicated that large lesions that include AI do not effect learned frequency discriminations in cats (Butler et al., 1957; Goldberg and Neff, 1961), but deficits have been shown in other behavioral paradigms (Meyer and Woolsey, 1952; Thompson, 1960). These studies suggest that the sequence and timing of stimuli, or perhaps the difficulty of the task, affect the necessary contribution of auditory cortical fields regarding this behavior. The behavioral task used in these experiments cannot be directly compared to those cited above, yet were considered by human observers to be quite difficult and to require vigilant attention in order to perform correctly. An alternative view is that the pairs of tone pips presented sequentially in the present study more closely match auditory pattern discrimination paradigms, which do show deficits following auditory cortical lesions (Diamond and Neff, 1957). A third point to bear in mind is that the limited data available suggest that thalamic areas with frequency-specific responses (ventral medial geniculate nucleus) are less affected by classical conditioning paradigms than the cerebral cortex, although the responses of the more broadly tuned neurons located in the magnocellular division of the medial geniculate are affected by classical conditioning paradigms (see Weinberger et al., 1990, for review). The cortical representations described in this report may be a reflection of tonotopically reorganized subcortical regions, partially reorganized subcortical and cortical areas, or resulting from reorganization only at the cortical level.

Regardless of the fact that we cannot identify the site(s) of reorganization, the differences in the response properties of AI neurons between trained and untrained monkeys do provide some insights into plausible neural mechanisms of frequency discrimination. Frequency discrimination abilities could potentially be based on either the rate of neuronal discharge, the temporal information encoded within the discharge, and/or the spatial location/distribution of the neurons responding to the different frequencies. Our finding of an increased spatial representation of behaviorally relevant frequencies suggests that the spatial distribution of central neurons at least contributes to this behavior. One simple hypothesis is that the separation in the spatial locations of appropriately tuned neurons could provide the necessary information to form the basis of the discrimination. Increases in the representation of a frequency range would increase the spatial resolution of those frequencies. The fact that the bandwidth of the tuned response was not significantly different at 40 dB above threshold suggests that this larger representation was maintained at higher intensities, for example, those used in the behavioral paradigm. However, the cortical area measure at CF described here is almost certainly an oversimplification. By combining the several islands of cortical representation into a single measure, we may have diluted the contributing portion. Studies in the cat have shown that several different parameters of both tuning and timing of the neural response are topographically organized along the isofrequency axis (Mendelson et al., 1988; Schreiner et al., 1988; Schreiner and Mendelson, 1990; Imig et al., 1991; Sutter and Schreiner, 1991), and preliminary data suggest that this is also the case in the owl monkey (Merzenich et al., 1991). It is plausible that only a subset of the total area of representation is contributing to the processing of frequency distinctions, while other parts of the representation are processing other stimulus parameters. Finally, the neural responses were not tested with respect to the their temporal representation of the stimuli. Studies in somatosensory cortex have shown that the occurrences of neural discharges are temporally more coherent in monkeys trained at a tactile frequency discrimination task when compared to unstimulated and passively stimulated monkeys (Recanzone et al., 1992d). The temporal coding of the stimuli could provide additional information regarding frequency.

The finding that Q10dB did increase significantly was unexpected, as the behavioral stimuli were not presented at these low intensities. The disparity between Q10dB and Q40dB suggests that these two parameters are not equivalent, consistent with observations in cat (Schreiner and Mendelson, 1990; Sutter and Schreiner, 1991). Q10dB measures have been reported to covary with latency, and both were observed to increase in the trained monkeys. The longer latencies in the trained monkeys may have resulted from increased neural processing at subcortical levels, or perhaps from a suppression of normally shorter-latency inputs that are involved in processing other aspects of the acoustic signal. These two parameters were not correlated with the behavioral performance, as the higher Q10dB (narrower bandwidths) and latency measures were not matched with the lowest thresholds.

Finally, both hemispheres were investigated in only one monkey. The hemisphere studied in each of the other trained monkeys was selected as the hemisphere contralateral to the ear nearest the speaker in the behavioral apparatus. In most cases, the monkey oriented its head either directly toward the speaker, in which case the ears were at an equivalent distance, or toward the food hopper, which would slightly favor one ear over the other. Both hemispheres of OM1 had similar areas of representations; however, there were significant differences between the right and left hemispheres with respect to both Q10dB and Q40dB. It is unclear at present how these binaural stimuli are processed by the two hemispheres.

Comparisons with earlier studies. The caveats discussed above do not detract from the findings that there is a significant improvement in behavioral performance at the frequency discrimination task with training that was significantly correlated with the spatial representation of the trained frequencies in primary auditory cortex. Improvements in performance at a perceptual or motor skill have been well documented for a variety of tasks (see Volkmann, 1858; James, 1890; Gibson, 1953; Fitts, 1964), including relatively simple sensory discrimination tasks in adult primates (Sinnott et al., 1985; Prosen et al., 1990; Recanzone et al., 1992a), and have been shown to continue for up to several years of continuous practice. The initial, rapid component of the improvement probably represents a "conceptual" learning of the task and development of the most appropriate strategies (see deJong, 1957; Crossman, 1959; Seibel, 1963; Recanzone et al., 1992a). This is supported by the large variability in the performance between these early sessions, and by the fact that it was not seen when the task was changed to different frequencies in well-trained monkeys (e.g., 2.5 kHz to 8 kHz in OM2).

The second stage of learning was characterized by an improvement in performance at each S2 frequency in a roughly sequential fashion, and an increase in the slope of the psycho-

metric function at threshold. This improvement has been hypothesized to result from an enhancement of the central representation of the appropriate stimulus parameters (Recanzone et al., 1992a,d). The psychophysical results derived in these monkeys are in agreement with this hypothesis. For example, there was little transference of the improvement in performance when untrained frequency ranges were tested in widely separated sessions, consistent with studies on absolute auditory detection thresholds (Zwislocki et al., 1958) or tactile frequency discrimination thresholds (Recanzone et al., 1992a). Improvements were only recorded when training was continuous for several days. The representations of these untrained frequencies were comparable to those of normal monkeys. An interesting observation was the apparent decrease in performance for decreasing S2 frequencies ($-\Delta F$ values) with training at increasing S2 frequencies ($+\Delta F$ values). The thresholds for $-\Delta F$ values have been reported to be slightly lower in humans but vary depending on the species for Old-World monkeys (Sinnott et al., 1987). In those experiments, the behavioral performances were reported following training at both increasing and decreasing ΔF tasks. In OM4, where several measures of $-\Delta F$ threshold were made, there were no cortical locations observed where the CF was centered in the range of these decreasing S2 frequencies. Finally, our psychophysical results suggest that there is little, if any, effect of stimulus level on discrimination performance. There was a slight decrease in threshold with increasing stimulus level, but this difference was not statistically significant. Increasing intensity can increase frequency discrimination ability in humans, but decreases performance in macaque monkeys (Sinnott et al., 1987). Our results suggest either that frequency discrimination performance is normally independent of stimulus level in this species, or that the performance at the trained stimulus level improved to match that for lower stimulus levels. The effects of stimulus amplitude on frequency discrimination in a naive animal of this species would provide further information on this issue.

The electrophysiological results are consistent with a variety of preparations in the somatosensory system (for reviews, see Merzenich et al., 1988, 1990; Kaas, 1991) and with the hypothesis that the functional organization of the cortex changes in parallel with perceptual changes (see Merzenich et al., 1988, 1990; Recanzone et al., 1991b,d; Recanzone and Merzenich, 1992). The finding that the large cortical representation of the behaviorally trained receptor surface was only seen in monkeys trained to discriminate those frequencies, and not in monkeys trained at an unrelated task, is consistent with results from the somatosensory system (Jenkins et al., 1990; Recanzone et al., 1992b–d). This representational plasticity occurred even though the monkeys in this study were acoustically stimulated for approximately 15–20 min during the 2 hr session each day, and were allowed to hear other auditory stimuli, some of which were presumably behaviorally significant, for the remaining 22 hr each day. Thus, it is not necessary to present the behaviorally important stimuli to the exclusion of all other stimuli for these effects to occur, and consistent with reports on the effects of classical conditioning on responses of auditory cortical neurons, behavioral relevance of the stimulus is necessary.

The demonstration that there is a larger CF representation in trained monkeys is not altogether unexpected given the frequency-specific changes in neural responses of other auditory cortical areas following classical conditioning paradigms in the cat (Diamond and Weinberger, 1986) and guinea pig (Bakin and Weinberger, 1990) and changes in the tonotopic representation of AI after restricted cortical lesions in the adult guinea pig and cat (Robertson and Irvine, 1989; Irvine et al., 1991). In contrast to the conditioning studies, there was no obvious change in the overall responsiveness of neurons representing the frequencies used in behavioral training, and there was no significant change in the tuning of these neurons at intensities well above threshold. Similarly, several months after restricted lesions in the cochlea, the overall activity of neurons in the expanded zone of CF representation in AI is similar to neurons in the adjacent, unaffected region. These differences may reflect differences in the comparison between short-term effects as in the conditioning studies and the long-term effects as in this and the lesion studies, in the anesthetic state of the animal, or in the differences between the conditioning and discrimination behavioral paradigms.

Several models advanced to account for cortical plasticity have proposed that synaptic efficacies are altered consequent to the temporal correlation of pre- and postsynaptic activity by a Hebbian mechanism (e.g., see Edelman, 1987; Merzenich et al., 1988, 1990; von der Malsburg and Singer, 1988; Palm, 1990; Singer, 1990; Weinberger et al., 1990). These mechanisms, although presented with respect to the cerebral cortex, could be equally effective for representational changes in subcortical areas. Computer models using these types of synaptic efficacy changes result in changes in topography similar to those observed in the somatosensory (Pearson et al., 1987; Grajski and Merzenich, 1990) and visual cortices (Miller et al., 1989). The present results are consistent with models of this class, as the increased area of representation and the increased sharpness of tuning could result from the strengthening of the synapses representing the behaviorally trained frequencies, and a weakening of synapses representing other, neighboring frequencies. The unaffected representations in the passive stimulation monkeys suggest that attention to the stimulus enhances the response, perhaps via neuromodulators such as ACh or norepinephrine (see Singer, 1990; Weinberger et al., 1990).

The inverse relationship between receptive field size and the cortical area of representation described in the visual and somatosensory systems (Hubel and Wiesel, 1974; Sur et al., 1980) is thought to arise by competitive interactions of excitatory and inhibitory inputs between cortical neurons comprising a horizontally oriented network (see Edelman, 1987; Merzenich et al., 1990). These local interactions between cortical neurons would also restrict the improvement to the behaviorally trained and immediately surrounding frequencies, and thus only limited effects on either the cortical representation or behavioral performance for other frequencies are expected. The decreasing performance at $-\Delta F$ frequencies observed in this study suggests that these horizontal influences degrade the representation of unused neighboring frequencies, perhaps by "borrowing" territory from the representation of those frequencies. The local interactions would also be expected to sharpen the temporal cohesiveness of the responses over a wider area, as has been demonstrated in the somatosensory cortex (Recanzone et al., 1992d), which would result in a stronger representation of the stimulus in both space and time that could then be transmitted to other areas for further processing.

Technical considerations. Several technical considerations must be remembered when attempting to correlate data derived by presenting single tone bursts in a closed sound system in an anesthetized monkey with behavioral measures performed using sequential tone bursts in the free field. The barbiturate anes-

thesia may have masked differential neuronal responses to behaviorally relevant stimuli. The effects of barbiturate anesthesia on tuning properties of cells in AI of the monkey are not well described; however, comparison of our data with that of others derived in awake animals (i.e., Funkenstein and Winter, 1973; Pelleg-Toiba and Wollberg, 1989) suggests little effect on measures of CF or temporal response properties of cortical neurons. In this study, all of the comparisons of the neural responses were made between monkeys anesthetized in a similar manner; thus, all anesthetic effects are constant and do not differentially affect one set of monkeys.

A second consideration is that the stimuli used in generating the cortical maps were shorter in duration (50 msec vs. 150 msec) and were presented through a closed sound system. The acoustic environment in the free field and the resonance properties of the external ear may have distorted the stimuli in the behavioral condition. The definition of response properties of AI neurons using a closed sound system has the advantage of better control of the stimulus intensity within and between individual experiments. Direct comparisons can also be made between these results and those of others. If the electrophysiological experiments were performed in the free field, idiosyncratic distortions of the stimulus could occur between individual monkeys due to variations in head size and location and pinna morphology. To match the acoustic environment of the behavioral apparatus in the electrophysiology experiment, it would have been necessary to perform both in the same apparatus, which was not possible. Finally, the technique we have used is appropriate to achieve the primary goal of the study, namely, to define alterations in the tonotopic organization of AI consequent to the behavioral training.

A third potential problem is in correlating the pooled data from normal monkeys with the behavioral results from other monkeys either early in training or when tested on untrained frequencies. This method increased the number of data points such that reasonable correlation analysis between physiological and behavioral data could be performed, but is based on the assumption that the physiological and behavioral measures are representative for all monkeys. The estimates of "untrained" behavioral performance are probably reasonable because the thresholds for sessions 5–10 in monkey OM1 were within 15% of the initial five sessions for OM2 at the 8 kHz task, as were the measures for sessions 5–10 for monkeys OM2 and OM3 in the 2.5 kHz task. These differences were not significantly different, and all of our data for both auditory and tactile frequency discrimination measures using this paradigm show low interindividual variability (see Recanzone et al., 1991, 1992a). There was also some variability in the cortical representations between normal monkeys, as has been seen previously in AI (Merzenich and Brugge, 1973; Imig et al., 1977; Aitken et al., 1986; see also Merzenich et al., 1975), consistent with interanimal variability described in primary somatosensory cortex (Merzenich et al., 1987). This variability was most evident in the large representation of frequencies in the range of 10 kHz in monkey N2 and for frequencies near 11 kHz in monkey CM1. This interanimal variability of frequency representations was reduced when considered from the point of view of the expanded representations in the trained monkeys, where cortical representations were invariably several times greater than the mean representations from untrained hemispheres. These idiosyncrasies of primate AI organization prompted us to utilize the strategy of using different comparison frequencies for each of the limited number of owl monkeys trained in the behavioral task. Obviously, if one had trained monkey N2 on a 10 kHz task, they may have been misled by the preexisting large representation of those frequencies. By the same logic, the probability of matching the frequency in the behavioral paradigm with the idiosyncratically overrepresented frequency in each of the studied individual monkeys is remote.

One potential inadequacy of the data is that frequency response areas, and hence the measures of Q10dB and Q40dB, were based on the single presentation of each of the 675 stimuli. Spontaneous activity occurring in the 50 msec time window of data collection could artificially broaden the response area, and only stimuli that evoke a response on several repeated trials might be consistently considered within the response area. However, this method has been shown to be a reliable measure of multiple-unit responses. The objective criteria used to define Q10dB and Q40dB take into account the rate of spontaneous activity as well as the consistency of the response latency, and have been argued to accurately reflect the tuning and timing properties of the neurons under study (see Schreiner and Mendelson, 1990; Sutter and Schreiner, 1991). The minimum latency estimate was based on all 15 levels at three frequencies, and therefore represents 45 separate measurements.

Concluding remarks. The demonstration that the extent of the cortical area of representation of a specific frequency range is correlated with the behaviorally measured discrimination performance represents a first approximation for determining the neural correlate of this perceptual judgement. The results of these experiments are consistent with hypotheses that auditory cortical representations are modifiable and that this plasticity reflects the acquisition of skills and behaviors throughout life. A key prediction from these hypotheses is that the changes in cortical representation should occur in parallel with the behavioral gains, and repeated measures of these representations should correlate with the repeated measures of behavioral performance. A second prediction is that inactivation of restricted areas of the central representation of these stimuli should result in restricted behavioral deficits. Finally, the effects of interrupting the normal release of putative neuromodulators during the period of acquisition of this behavioral improvement could begin to reveal the pharmacology of this representational plasticity as has been done in other systems (i.e, Kasamatsu and Pettigrew, 1979; Bear and Singer, 1986). Such studies would provide key evidence in extending these hypotheses and models of cortical function.

References

Aitkin LM, Merzenich MM, Irvine DRF, Clarey JC, Nelson JE (1986) Frequency representation in auditory cortex of the common marmoset (*Callithrix jacchus jacchus*). J Comp Neurol 252:175–185.

Ashe JH, McKenna TS, Weinberger NM (1989) Cholinergic modulation of frequency receptive fields in auditory cortex. II. Frequency-specific effects of anticholinesterases provide evidence for a modulatory action of endogenous ACh. Synapse 4:44–54.

Bakin JS, Weinberger NM (1990) Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig. Brain Res 536:271–286.

Bear MF, Singer W (1986) Modulation of visual cortical plasticity by acetylcholine and noradrenaline. Nature 320:172–176.

Beecher MD (1974) Hearing in the owl monkey (*Aotus trivirgatus*). I. Auditory sensitivity. J Comp Physiol Psychol 86:898–901.

Butler RA, Diamond IT, Neff WD (1957) Role of auditory cortex in discrimination of changes in frequency. J Neurophysiol 20:108–120.

Craig JC (1988) The role of experience in tactual pattern perception: a preliminary report. Int J Rehabil Res 11:167–171.

Crossman ERFW (1959) A theory of the acquisition of speed skill. Ergonomics 2:153–166.
deJong JR (1957) The effects of increasing skill on cycle time and its consequences for time standards. Ergonomics 1:51–60.
Diamond DM, Weinberger NM (1986) Classical conditioning rapidly induces specific changes in frequency receptive fields of single neurons in secondary and ventral ectosylvian auditory cortical fields. Brain Res 372:357–360.
Diamond IT, Neff WD (1957) Ablation of temporal cortex and discrimination of auditory patterns. J Neurophysiol 20:300–315.
Disterhoft JF, Stuart DK (1976) Trial sequence of changed unit activity in auditory system of alert rat during conditioned response acquisition and extinction. J Neurophysiol 39:266–281.
Edelman GM (1987) Neuronal Darwinism: the theory of neuronal group selection. New York: Basic.
Fitts PM (1964) Perceptual-motor skill learning. In: Categories of human learning (Melton AW, ed), pp 243–285. New York: Academic.
Funkenstein HH, Winter P (1973) Responses to acoustic stimuli of units in the auditory cortex of awake squirrel monkeys. Exp Brain Res 18:464–488.
Gibson EJ (1953) Improvement in perceptual judgments as a function of controlled practice or training. Psychol Bull 50:401–431.
Goldberg JM, Neff WD (1961) Frequency discrimination after bilateral ablation of cortical auditory areas. J Neurophysiol 24:119–128.
Grajski KA, Merzenich MM (1990) Hebb-type dynamics is sufficient to account for the inverse magnification rule in cortical somatotopy. Neural Comput 2:71–84.
Green DM, Swets JA (1966) Signal detection theory. New York: Wiley.
Hubel DH, Wiesel TN (1974) Uniformity of monkey striate cortex: a parallel relationship between field size, scatter and magnification factor. J Comp Neurol 158:295–302.
Imig TJ, Rugero MA, Kitzes LM, Javel E, Brugge JF (1977) Organization of auditory cortex in the owl monkey (*Aotus trivirgatus*). J Comp Neurol 171:111–128.
Imig TJ, Irons A, Samson FR (1991) Single-unit selectivity to azimuthal direction and sound pressure level of noise bursts in cat high-frequency primary auditory cortex. J Neurophysiol 63:1448–1466.
Irvine DRF, Rajan R, Wize LZ, Heil P (1991) Reorganization in auditory cortex of adult cats with unilateral restricted cochlear lesions. Soc Neurosci Abstr 17:1485.
James W (1890) The principles of psychology, Vol I. New York: Dover.
Jenkins WM, Merzenich MM, Ochs MT, Allard T, Guíc-Robles E (1990) Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. J Neurophysiol 63:82–104.
Kaas JH (1991) Plasticity of sensory and motor maps in adult mammals. Annu Rev Neurosci 14:137–167.
Kasamatsu T, Pettigrew JD (1979) Preservation of binocularity after monocular deprivation in the striate cortex of kittens treated with 6-hydroxydopamine. J Comp Neurol 185:139–162.
Kitzes LM, Farley GR, Starr A (1978) Modulation of auditory cortex unit activity during the performance of a conditioned response. Exp Neurol 62:678–697.
McKenna TM, Ashe JH, Weinberger NM (1989) Cholinergic modulation of frequency receptive fields in auditory cortex. I. Frequency-specific effects of muscarinic agonists. Synapse 4:30–43.
Mendelson JR, Schreiner CE, Grasse K, Sutter ML (1988) Spatial distribution of responses to FM sweeps in cat primary auditory cortex. Abstr Assoc Res Otolaryngol 11:36.
Merzenich MM, Brugge JF (1973) Representation of the cochlear partition on the superior temporal plane of the macaque monkey. Brain Res 50:275–296.
Merzenich MM, Knight PL, Roth GL (1975) Representation of cochlea within primary auditory cortex in the cat. J Neurophysiol 38:231–249.
Merzenich MM, Nelson RJ, Kaas JH, Stryker MP, Jenkins WM, Zook JM, Cynader MS, Schoppmann A (1987) Variability in hand surface representations in areas 3b and 1 in adult owl and squirrel monkeys. J Comp Neurol 258:281–396.
Merzenich MM, Recanzone G, Jenkins WM, Allard TT, Nudo RJ (1988) Cortical representational plasticity. In: Neurobiology of neocortex (Rakic P, Singer W, eds), pp 41–67. New York: Wiley.
Merzenich MM, Recanzone GH, Jenkins WM, Nudo RJ (1990) How the brain functionally rewires itself. In: Natural and artificial parallel computations (Arbib M, Robinson JA, eds), pp 177–210. Cambridge, MA: MIT Press.
Merzenich MM, Schreiner CE, Recanzone GH, Beitel RE, Sutter ML (1991) Topographic organization of cortical field AI in the owl monkey *Aotus trivirgatus*. Abstr Assoc Res Otolaryngol 14:44.
Metherate R, Weinberger NM (1990) Cholinergic modulation of responses to single tones produces tone-specific receptive field alterations in cat auditory cortex. Synapse 6:133–145.
Meyer DR, Woolsey CN (1952) Effects of localized cortical destruction on auditory discriminative conditioning in cat. J Neurophysiol 15: 149–162.
Miller KD, Keller JB, Stryker MP (1989) Ocular dominance column development: analysis and simulation. Science 245:605–615.
Olds J, Disterhoft JF, Segal M, Kornblith CL, Hirsh R (1972) Learning centers of rat brain mapped by measuring latencies of conditioned unit responses. J Neurophysiol 35:202–219.
Palm G (1990) Cell assemblies as a guideline for brain research. Concepts Neurosci 1:133–147.
Pearson JC, Finkel LH, Edelman GM (1987) Plasticity in the organization of adult cerebral cortical maps: a computer simulation based on neuronal group selection. J Neurosci 7:4209–4223.
Pelleg-Toiba R, Wollberg Z (1989) Tuning properties of auditory cortex cells in the awake squirrel monkey. Exp Brain Res 74:353–364.
Prosen CA, Moody DB, Sommers MS, Stebbins WC (1990) Frequency discrimination in the monkey. J Acoust Soc Am 88:2152–2158.
Recanzone GH, Merzenich MM (1992) Alterations of the functional organization of primary somatosensory cortex following intracortical microstimulation or behavioral training. In: Memory: organization and locus of change (Squire LM, Weinberger NM, Lynch G, McGaugh JL, eds), pp 217–238. New York: Oxford UP.
Recanzone GH, Jenkins WM, Hradek GT, Merzenich MM (1991) A behavioral frequency discrimination paradigm for use in adult primates. Behav Res Methods Instrum Comput 23:357–369.
Recanzone GH, Jenkins WM, Hradek GT, Merzenich MM (1992a) Progressive improvements in discriminative abilities in adult owl monkeys performing a tactile frequency discrimination task. J Neurophysiol 67:1015–1030.
Recanzone GH, Merzenich MM, Jenkins WM, Grajski KA, Dinse HR (1992b) Topographic reorganization of the hand representation in cortical area 3b of owl monkeys trained in a frequency discrimination task. J Neurophysiol 67:1031–1056.
Recanzone GH, Merzenich MM, Jenkins WM (1992c) Frequency discrimination training engaging a restricted skin surface results in an emergence of a cutaneous response zone in cortical area 3a. J Neurophysiol 67:1057–1070.
Recanzone GH, Merzenich MM, Schreiner CE (1992d) Changes in the distributed temporal response properties of SI cortical neurons reflect improvements in performance on a temporally-based tactile discrimination task. J Neurophysiol 67:1071–1091.
Robertson D, Irvine DRF (1989) Plasticity of frequency organization in auditory cortex of guinea pigs with partial unilateral deafness. J Comp Neurol 282:456–471.
Ryugo DK, Weinberger NM (1978) Differential plasticity of morphologically distinct populations in the medial geniculate body of the cat during classical conditioning. Behav Biol 22:275–301.
Schreiner CE, Mendelson JR (1990) Functional topography of cat primary auditory cortex. Distribution of integrated excitation. J Neurophysiol 64:1442–1459.
Schreiner CE, Mendelson JR, Grasse K, Sutter ML (1988) Spatial distribution of basic response properties in cat primary auditory cortex. Abstr Assoc Res Otolaryngol 11:36.
Seibel R (1963) Discrimination reaction time for a 1,023-alternative task. J Exp Psychol 66:215–226.
Singer W (1990) Search for coherence: a basic principle of cortical self-organization. Concepts Neurosci 1:1–26.
Sinnott JM, Petersen MR, Hopp SL (1985) Frequency and intensity discrimination in humans and monkeys. J Acoust Soc Am 78:1977–1985.
Sinnott JM, Owren MJ, Petersen MR (1987) Auditory frequency discrimination in primates: species differences (*Cercopithecus, Macaca, Homo*). J Comp Psychol 101:126–131.
Sur M, Merzenich MM, Kaas JH (1980) Magnification, receptive-field area, and "hypercolumn" size in areas 3b and 1 of somatosensory cortex in owl monkeys. J Neurophysiol 44:295–311.
Sutter ML, Schreiner CE (1991) Physiology and topography of neurons

with multipeaked tuning curves in cat primary auditory cortex. J Neurophysiol 65:1207–1226.
Thompson RF (1960) Function of auditory cortex of cat in frequency discrimination. J Neurophysiol 23:321–334.
Volkman AW (1858) Über den Einfluss der Übung. Leipzig Ber Math-Phys Classe 10:38–69.
von der Malsburg C, Singer W (1988) Principles of cortical network organization. In: Neurobiology of neocortex (Rakic P, Singer W, eds), pp 69–99. New York: Wiley.
Weinberger NM, Ashe JH, Metherate R, McKenna TM, Diamond DM, Bakin J (1990) Returning auditory cortex by learning: a preliminary model of receptive field plasticity. Concepts Neurosci 1:91–132.
Wier CC, Jesteadt W, Green DM (1977) Frequency discrimination as a function of frequency and sensation level. J Acoust Soc Am 61: 178–184.
Wong-Riley MTT, Merzenich MM, Leake PS (1978) Changes in endogenous enzymatic reactivity to DAB induced by neuronal inactivity. Brain Res 141:185–192.
Zwislocki J, Maire F, Feldman AS, Rubin H (1958) On the effect of practice and motivation on the threshold of audibility. J Acoust Soc Am 30:254–262.

Systems/Circuits

Prefrontal Cortex-Driven Dopamine Signals in the Striatum Show Unique Spatial and Pharmacological Properties

Martín F. Adrover,[1*] Jung Hoon Shin,[1*] Cesar Quiroz,[2] Sergi Ferré,[2] Julia C. Lemos,[1] and Veronica A. Alvarez[1,3]

[1]Laboratory on Neurobiology of Compulsive Behavior, Intramural Research Program, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, Maryland 20892, [2]Integrative Neurobiology Section, National Institute on Drug Abuse, Intramural Research Program, National Institutes of Health, Baltimore, Maryland 21224, and [3]Center on Compulsive Behaviors, National Institutes of Health, Bethesda, Maryland 20892

Dopamine (DA) signals in the striatum are critical for a variety of vital processes, including motivation, motor learning, and reinforcement learning. Striatal DA signals can be evoked by direct activation of inputs from midbrain DA neurons (DANs) as well as cortical and thalamic inputs to the striatum. In this study, we show that *in vivo* optogenetic stimulation of prelimbic (PrL) and infralimbic (IL) cortical afferents to the striatum triggers an increase in extracellular DA concentration, which coincides with elevation of striatal acetylcholine (ACh) levels. This increase is blocked by a nicotinic ACh receptor (nAChR) antagonist. Using single or dual optogenetic stimulation in brain slices from male and female mice, we compared the properties of these PrL/IL-evoked DA signals with those evoked by stimulation from midbrain DAN axonal projections. PrL/IL-evoked DA signals are undistinguishable from DAN evoked DA signals in their amplitudes and electrochemical properties. However, PrL/IL-evoked DA signals are spatially restricted and preferentially recorded in the dorsomedial striatum. PrL/IL-evoked DA signals also differ in their pharmacological properties, requiring activation of glutamate and nicotinic ACh receptors. Thus, both *in vivo* and *in vitro* results indicate that cortical evoked DA signals rely on recruitment of cholinergic interneurons, which renders DA signals less able to summate during trains of stimulation and more sensitive to both cholinergic drugs and temperature. In conclusion, cortical and midbrain inputs to the striatum evoke DA signals with unique spatial and pharmacological properties that likely shape their functional roles and behavioral relevance.

Key words: DA release; dorsomedial striatum; fast-scan cyclic voltammetry; optogenetics; PFC

Significance Statement

Dopamine signals in the striatum play a critical role in basal ganglia function, such as reinforcement and motor learning. Different afferents to the striatum can trigger dopamine signals, but their release properties are not well understood. Further, these input-specific dopamine signals have only been studied in separate animals. Here we show that optogenetic stimulation of cortical glutamatergic afferents to the striatum triggers dopamine signals both *in vivo* and *in vitro*. These afferents engage cholinergic interneurons, which drive dopamine release from dopamine neuron axons by activation of nicotinic acetylcholine receptors. We also show that cortically evoked dopamine signals have other unique properties, including spatial restriction and sensitivity to temperature changes than dopamine signals evoked by stimulation of midbrain dopamine neuron axons.

Received May 27, 2020; revised July 22, 2020; accepted Aug. 17, 2020.

Author contributions: M.F.A., J.H.S., and V.A.A. designed research; M.F.A., J.H.S., C.Q., S.F., and J.C.L. performed research; M.F.A., J.H.S., C.Q., S.F., and J.C.L. analyzed data; M.F.A., J.H.S., and V.A.A. wrote the first draft of the paper; M.F.A., J.H.S., C.Q., S.F., J.C.L., and V.A.A. edited the paper; M.F.A., J.H.S., and V.A.A. wrote the paper.

*M.F.A. and J.H.S. contributed equally to this work.

This work was supported by the Intramural Programs of National Institute on Alcohol Abuse and Alcoholism, National Institute of Neurological Disorders and Stroke (ZIA-AA000421), and National Institute on Drug Abuse. We thank Roland Bock (National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health) for development of the VIGOR acquisition and analysis software; Drs. Karl Deisseroth (Stanford University) and Ed Boyden for providing the channelrhodopsin-2 and ChrimsonR constructs, respectively; and members of the A.A.V. laboratory for valuable comments on the manuscript.

M.F. Adrover's present address: Instituto de Investigaciones en Ingeniería Genética y Biología Molecular, CONICET, Buenos Aires, C1428ADN, Argentina.

J.C. Lemos' present address: Department of Neuroscience, University of Minnesota, Minneapolis, Minnesota 55455.

The authors declare no competing financial interests.

Correspondence should be addressed to Veronica A. Alvarez at alvarezva@mail.nih.gov.

https://doi.org/10.1523/JNEUROSCI.1327-20.2020

Copyright © 2020 the authors

Introduction

The striatum receives dense innervation from midbrain DA neurons (DANs), which are the main source of DA in the striatum. DA plays a critical neuromodulatory role in regulating striatal circuitry and function (Surmeier et al., 2007; Gerfen and Surmeier, 2011; Burke et al., 2017). Disruptions in striatal DA levels are associated with many neurologic and psychiatric disorders, such as Parkinson's disease and substance abuse disorder (Gerfen, 2000; Luscher et al., 2020). DA is released from a fraction of the varicosities distributed along DAN axons, which contain specialized active zone-like sites (Sulzer et al., 2016; Liu et al., 2018; Liu and Kaeser, 2019). The axonal arborizations of DANs ramify extensively within the striatum, and a single dopamine axon can spread over a significant area (~3% in average)

of the striatum (Matsuda et al., 2009). DA released from axonal varicosities generates a rapid increase in local extracellular DA concentration. This extracellular increase leads to the activation of multiple types of DA receptors, which are localized in dendrites, somas, and presynaptic terminals. Ultimately, this impacts the network activity of the striatum and drives behavior output (Tritsch and Sabatini, 2012; Chuhma et al., 2014; Mamaligas et al., 2016; Burke et al., 2017; Shin et al., 2017; Lahiri and Bevan, 2020). This striatal DA signal is known to be triggered by two mechanisms. One mechanism involves action potential firing initiated at midbrain DAN somas, which propagates through dense axonal arborizations to reach active zone-like release sites in the striatum (Matsuda et al., 2009; Liu et al., 2018). In the other mechanism, DAN axons in the striatum are locally activated, independent of the action potential firing at DAN somas. This mode of DA release requires the activation of nAChRs expressed on DAN axons and synchronized activation of cholinergic interneurons (CINs), which is thought to give rise to local release of ACh (Cachope et al., 2012; Threlfell et al., 2012; Wang et al., 2014; Shin et al., 2015, 2017).

This local intrastriatal trigger of DA release has only recently been demonstrated and is gaining attention as it represents a newly discovered means for striatal DA modulation/transmission. However, little is known about the differential/unique biophysical and pharmacological properties of locally evoked versus DAN-evoked DA transmission. Recent *in vitro* studies using optogenetic stimulation and fast-scan cyclic voltammetry (FSCV) show that local DA signals can be triggered in the striatum by stimulation of glutamatergic inputs from thalamus (Threlfell et al., 2012; Kosillo et al., 2016; Johnson et al., 2017; Cover et al., 2019), motor cortex (Kosillo et al., 2016), or PFC (Mateo et al., 2017). These findings agree with previous literature showing *in vivo* DA signals in the striatum evoked by stimulation of PFC (Taber and Fibiger, 1993; Quiroz et al., 2016; Hill et al., 2018), hippocampus (Tritschler et al., 2018), or amygdala (Floresco et al., 1998). Furthermore, *in vivo* intrastriatal administration of glutamate and ACh was also shown to cause increases in extracellular DA concentration in the rat striatum (Giorgieff et al., 1976, 1977; Leviel et al., 1990; Shimizu et al., 1990). Together, this evidence supports an intrastriatal mechanism for DA signal generation that can be initiated by excitatory inputs to the striatum that require glutamate and ACh. While cortically evoked DA signals in the striatum have been reported, identifying the unique pharmacological and basic properties of these input-specific signals will allow us to selectively manipulate and target them, leading to a better understanding of their functional significance. In this study, we tackle this gap in knowledge using both *in vivo* microdialysis and *in vitro* FSCV with optogenetic stimulation. Particularly, we set up a novel approach using dual optogenetic stimulation in the same brain slices to input-specific evoke DA release and compare the properties of these DA signals in the striatum.

Materials and Methods

Animals. All animals used in the study were maintained in accordance with the guidelines of the National Institutes of Health Animal Care, and the animal research procedures were approved by the National Institute on Alcohol Abuse and Alcoholism Animal Care and Use Committee for mice and by the National Institute on Drug Abuse Intramural Research Program Animal Care and Use Committee for rats. Except for the *in vivo* microdialysis experiments, all experiments were conducted using male and female mice of C57BL/6 background. Heterozygote B6.SJL-*Slc6a3*$^{tm1.1(cre)Bkmn}$/J mice (Backman et al., 2006) (The Jackson Laboratory, 006660) referred to as DAT$^{IRES\text{-}Cre+}$ mice and negative DAT$^{IRES\text{-}Cre-}$ littermates considered as WT mice were used. To fluorescently label CINs, homozygote B6N.129S6(B6)-*Chat*$^{tm2(cre)Lowl}$/J mice (Rossi et al., 2011) (The Jackson Laboratory, 018957) were crossed with homozygote B6.Cg-*Gt(ROSA)26Sor*$^{tm14(CAG\text{-}tdTomato)Hze}$/J mice (Madisen et al., 2010) (The Jackson Laboratory, 007914), and the progeny is referred to as CIN-tdTomato mice. For the microdialysis experiments, male Sprague Dawley albino rats (Charles River Laboratories) were used. Rats were housed individually for the first week after intracranial injection, until suture removal, after which the rats were housed 2 per cage. All animals were housed on a 12 h light/dark cycle (06:30 to 18:30 light) with food and water *ad libitum*.

Surgery and stereotaxic injection of AAV-ChR2 vectors. Injections were conducted as described previously (Adrover et al., 2014). Briefly, mice (5-6 weeks old) were anesthetized by inhalation of isoflurane-oxygen mixture and placed in a stereotaxic frame (David Kopf). Adeno-associated virus (AAV) vectors with Cre recombinase-dependent expression of Channelrhodopsin2 (ChR2) protein, AAV5-EF1a-DIO-hChR2(H134R)-EYFP (4×10^{12} IU/ml), were bilaterally injected into the VTA/SNc (AP: −3.30, ML: ±0.60, DV: −4.50) of DAT$^{IRES\text{-}Cre+}$ mice. The AAV vectors AAV5-CaMKII-hChR2(H134R)-EYFP (4×10^{12} IU/ml) or AAV5-Syn-ChrimsonR-tdTomato (1.7×10^{13} IU/ml), were injected in the PFC (PrL cortex, AP: 2.10, ML: ±0.35, DV: −2.30; IL cortex, AP: 1.90, ML: ±0.30, DV: −3.20) of negative littermates or DAT$^{IRES\text{-}Cre+}$ mice. All stereotaxic coordinates were from bregma (in mm) according to the mouse atlas by Franklin and Paxinos (2007); 300-500 nl for VTA/SNc and 100-200 nl for PFC were injected at a flow rate of 100 nl/min. Recordings were made after a minimum of 4 weeks of incubation. For mice injected in both VTA/SNc (DIO-ChR2-EYFP) and PFC (ChrimsonR-tdTomato), recordings were performed after a minimum of 8 weeks of incubation. Viral vectors were purchased from Gene Therapy Center Vector Core at University of North Carolina and Penn Vector Core at University of Pennsylvania.

Slice preparation. Mice were anesthetized and rapidly decapitated. Brains were quickly removed, mounted, and sliced using a vibratome (VT-1200S, Leica Microsystems) in an ice-cold cutting solution containing the following (in mM): 225 sucrose, 13.9 NaCl, 26.2 $NaHCO_3$, 1 NaH_2PO_4, 1.25 glucose, 2.5 KCl, 0.1 $CaCl_2$, 4.9 $MgCl_2$, and 3 kynurenic acid. The sagittal slices (240 μm) were incubated for 20 min at 33°C in ACSF containing the following (in mM): 124 NaCl, 1 NaH_2PO_4, 2.5 KCl, 1.3 $MgCl_2$, 2.5 $CaCl_2$, 20 glucose, 26.2 $NaHCO_3$, and 0.4 ascorbic acid, and kept in the dark at room temperature before use. Recording chamber was perfused at 2 ml/min with ACSF heated at 32°C using an inline heater (Harvard Apparatus).

FSCV and amperometry. FSCV was performed in the dorsal striatum. Carbon-fiber electrodes (CFEs) were prepared with a cylindrical carbon-fiber (7 μm diameter, ~150 μm of exposed fiber) inserted into a glass pipette. Before use, the CFEs were conditioned with an 8-ms-long triangular voltage ramp (−0.4 to 1.2 and back to −0.4 V vs Ag/AgCl reference at 400 V/s) delivered every 15 ms. CFEs showing current >1.8 μA or <1.0 μA in response to the voltage ramp at ~0.6 V were discarded. During the recording, the CFEs were held at −0.4 V versus Ag/AgCl, and the same triangular voltage ramp was delivered every 100 ms. DA transients were evoked by electrical or optical single pulse, or 5 pulses at 20 Hz stimulations. Using the same CFE and location, DA signals were evoked by alternating electrical and optical stimulations in some experiments. These data were combined with results from other experiments where either electrical or optical stimulation was used to evoke DA signals. For electrical stimulation, a glass pipette filled with ACSF was placed near the tip of the carbon fiber and a rectangular pulse (0.2 ms, 100 μA) was applied every 2 min. For optogenetic stimulation, a fiber-optic (200 μm diameter, 0.22 NA, ThorLabs) connected to a blue LED (470 nm, 1.8 mW of maximal output power measured at the tip of the fiber-optics, ThorLabs) was placed over the carbon fiber and light pulses (0.6 ms) were delivered every 2 min. For input-output curves, the widths of light pulse were 0.1, 0.2, 0.5, 1, 2, and 5 ms. For dual optogenetic recordings, two fiber-optics connected to a purple LED (420 nm, 3.0 mW, ThorLabs) and an orange LED (590 nm, 0.7 mW, ThorLabs), respectively, were placed over the carbon fiber. Light pulses (0.6 ms for 420 nm and 0.6-2 ms for 590 nm) were delivered in an alternating pattern every 2 min. Data were collected with a retrofit headstage (CB-7B/

EC with 5 MΩ resistor) using a Multiclamp 700B amplifier (Molecular Devices) after low-pass filter at 3 kHz and digitized at 100 kHz using a DA board (NI USB-6229 BNC, National Instruments). Data acquisition and analysis were performed using a custom-written software, VIGOR, in Igor Pro (Wavemetrics) using mafPC (courtesy of MA Xu-Friedman). The current peak amplitudes of the evoked DA transients were converted to DA concentration according to the postexperimental calibration using 1-3 μM DA. Amperometric recordings were performed using the same carbon-fiber electrodes held at −0.4 V versus Ag/AgCl and a 30-s-long step to 0.6 V was applied. Single pulse or 5 pulses at 20 Hz stimulation were delivered at 20 s after switching to 0.6 V. Since we did not find major differences between PrL and IL in evoking DA transients, we combined the data obtained from two groups to represent DA transients evoked by PFC inputs (PFC-oDA).

Cell-attached recordings. Striatal CINs from CIN-TdTomato mice injected with ChR2-EYFP in the PFC were identified by fluorescence and confirmed by their characteristic spontaneous firing pattern. Cell-attached recordings were performed from CINs in the striatum using glass pipette electrodes with a resistance of ~3-4 MΩ, filled with an internal solution containing the following (in mM): 120 cesium methanesulfonate, 20 CsCl, 10 HEPES, 0.2 EGTA, 10 sodium phosphocreatine, 4 Na2-ATP, and 0.4 Na-GTP, pH, 7.25 (290-310 mOsm), at a holding potential of 0 mV. To estimate the strength of connectivity to CINs, the lowest light intensity needed to reliably evoke action potentials in CINs was determined, ranging from 0.4 mW for "high" connectivity to 2.1 mW for "low" connectivity. The data were collected using a Multiclamp 700B amplifier after low-pass filter at 1 kHz and digitized at 5 kHz, using pClamp10 software (Molecular Devices). Spike fidelity was calculated as percentages of the number of AP evoked from five trials for both single pulse and trains stimulation.

In vivo *microdialysis and optogenetic stimulation of the cortical inputs.* Rats (80-90 g) received unilateral intracranial injection of AAV-CaMKIIa-hChR2(H134R)-EYFP (titer: 10^{12} IU/ml; Gene Therapy Center Vector Core at University of North Carolina) in the PrL and IL cortex (AP: 3.0 mm, ML: 0.5 mm, DV: −3.5 and −5 mm with respect to bregma). Two injection sites per hemisphere were used, and 300 nl of viral vector solution was delivered per site via a 105-μm-thick silica tubing injector coupled directly to a 1 μl syringe driven by an infusion pump (rate = 50 nl/min) during 10 min. The injector was left in place for an additional 10 min following each injection to allow for diffusion of the suspension. Ten weeks after viral vector injection, rats (350-400 g) underwent surgery for probe implantation according to previously published procedures (Quiroz et al., 2016). Briefly, a modified microdialysis probe with an embedded light-guiding optic fiber was implanted into the NAc shell (AP: 1.2 mm, ML: 0.5 mm, DV: −8.0 mm with respect to bregma). The probe was fixed to the skull with a stainless-steel screw and glass-ionomer dental cement. All surgical procedures were performed under anesthesia with 3 ml/kg of Equithesin (4.44 × *g* of chloral-hydrate, 0.972 × *g* of Na pentobarbital, 2.124 × *g* of $MgSO_4$, 44.4 ml of propylene glycol, 12 ml of ethanol, and distilled H_2O up to 100 ml of final solution; National Institute on Drug Abuse Pharmacy). To build the microdialysis probe with embedded optic fiber, the tip of an optic fiber (105 μm diameter core, 0.22 NA) was sculpted into a conical shape using a Flaming-Brown pipette puller fitted with a custom platinum heating filament (Sutter Instruments) to allow for a larger area of stimulation. The conical optic fiber tip was embedded inside the microdialysis probe and implanted. Microdialysis experiments were performed in freely moving rats 24 h after probe implantation. Optical fiber was coupled to a 473 nm solid-state laser module and light stimulation was driven by a Grass S88 stimulator. Light stimulation was delivered for 20 min as trains of light pulses (2 ms pulse duration; at 100 Hz for 160 ms; trains repeats once per second; intensity = 5-8 mW at probe tip). ACSF containing (in mM) 144 NaCl, 4.8 KCl, 1.7 $CaCl_2$, and 1.2 $MgCl_2$ was pumped through the optogenetic-microdialysis probe (rate = 1.25 μl/min). After a washout period of 90 min, dialysate samples were collected at 20 min intervals. After 80 min of baseline sampling, optogenetic stimulation was applied for 20 min, and samples were collected for 80 additional minutes after the end of the stimulation. Samples were split and analyzed separately for glutamate, DA, and ACh content. ACh and glutamate contents were measured by HPLC coupled to an ACh oxidase and glutamate oxidase enzyme reactors, respectively, and electrochemical detection (Eicom). DA was measured by HPLC coupled with a coulometric detector (5200a Coulochem III, ESA). At the end of the microdialysis experiment, animals were deeply anesthetized with Equithesin and perfused transcardially with 0.1 M PBS, followed by 4% formaldehyde in 0.1 M PBS, pH 7.4. Brains were postfixed in the same fixative for 2 h and immersed in 20% sucrose/0.1 M PBS, pH 7.4, solution for 48 h at 4°C. Forty-micron-thick coronal sections were cut in a Leica Microsystems CM3050S cryostat at −20°C, collected in PBS, and stored in antifreeze-buffered solution (20% ethylene glycol, 10% glycerol, and 10% sucrose in PBS) at −80°C until processing. Sections were then evaluated for localization of implanted probes and ChR2-EYFP expression. Wide-field images were acquired with a Typhoon laser scanner (GE Healthcare). Confocal fluorescence microscopy images were acquired with an Examiner Z1 microscope (Carl Zeiss) fitted with a confocal laser module (LSM-710, Carl Zeiss).

Drugs. Dihydro-β-erythroidine hydrobromide (DHβE) was purchased from Tocris. Kynurenic acid (sodium salt), NBQX, and 3-((*R*)-2-carboxypiperazin-4-yl)-propyl-1-phosphonic acid (CPP) were purchased from Abcam. All other chemicals were purchased from Sigma Millipore.

Statistical analysis. Statistical analysis was performed with Prism (GraphPad). One-sample *t* test, two-tailed paired *t* test, repeated-measures one-way ANOVA with or without mixed-effects models, or two-way ANOVA was used as specified. Tukey's or Dunnett's multiple comparisons test was used for *post hoc* analysis as specified. The number of experiments, *n*, was expressed as the number of slices or cells/the number of mice or number of rats for the *in vivo* microdialysis experiments.

Results

In vivo stimulation of PrL/IL inputs increases striatal ACh and DA concentration, and the increase in DA requires nAChR activation

Published work *in vivo* shows that optogenetic stimulation of cortical inputs to the striatum can evoke DA release (Quiroz et al., 2016). However, this study provides no evidence that activation of these cortical inputs recruits striatal CINs or that the cortically evoked DA release *in vivo* requires activation of striatal nAChRs. To assess these matters, we used *in vivo* microdialysis combined with optogenetic stimulation, as previously described (Quiroz et al., 2016), and measured extracellular levels of ACh and DA in response to optogenetic stimulation of cortical inputs. A modified microdialysis probe with an embedded light-guiding optic fiber was implanted into the striatum of rats expressing ChR2-EYFP in the PrL/IL cortex (Fig. 1*A*,*B*). On the experiment day, dialysate samples were collected at baseline for 80 min before stimulation started. Optogenetic stimulation of cortical axons within the striatum was delivered around the microdialysis probe as trains of light pulses (16 pulses at 100 Hz every 1 s for 20 min). Samples were collected for an additional 80 min after stimulation and were split and analyzed separately for DA and ACh content or DA and glutamate. This long train of *in vivo* optogenetic stimulation of PrL/IL cortical axon fibers within the striatum produced an increase in extracellular concentration of DA in the striatal dialysate (185.1 ± 44.0% of baseline, $n = 7$; $F = 5.27$, $p = 0.0002$, repeated-measures one-way ANOVA; $p = 0.0002$ for 60 min vs 80 min; Dunnett's multiple comparisons test; Fig. 1*C*), which coincided with an increase of extracellular ACh concentration (196.7 ± 36.6% of baseline, $n = 7$; $F = 3.00$, $p = 0.0002$, repeated-measures one-way ANOVA; $p = 0.004$ for 60 min vs 80 min; Dunnett's multiple comparisons test; Fig. 1*D*). These *in vivo* microdialysis findings support the hypothesis that cortical stimulation recruits CINs. In addition, when the *in vivo* optogenetic stimulation was performed in constant perfusion of

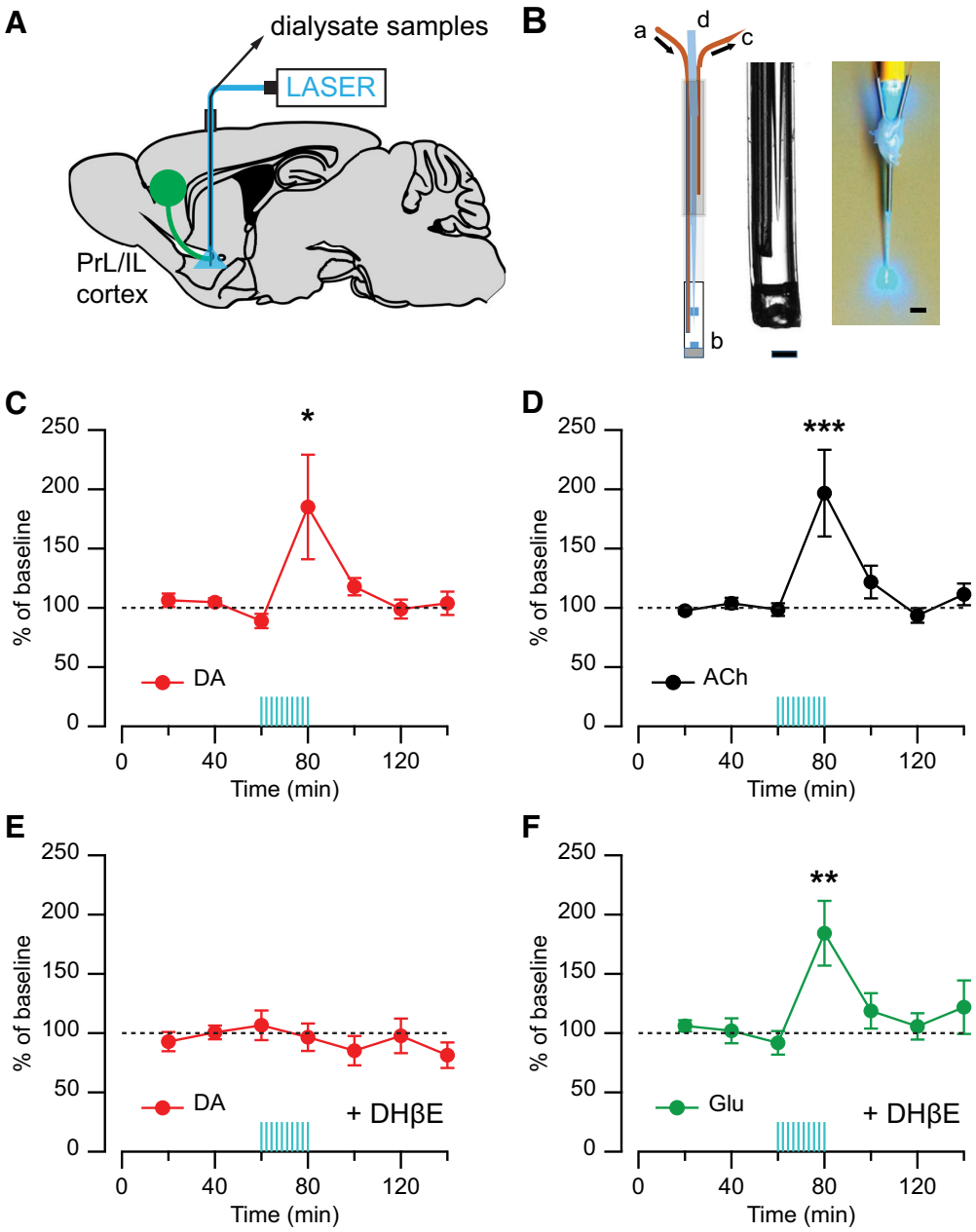


Figure 1. Effect of the nAChR antagonist DHβE on the extracellular DA level evoked by local optogenetic stimulation of PFC fibers *in vivo*. ***A***, Diagram showing the injection of ChR2-EYFP in the PrL and IL cortices (green circle) and projection to the striatum with the microdialysis-optogenetics probe in a sagittal brain view. ***B***, Left, Optogenetic-microdialysis probe schematics: a, liquid inlet; b, dialysis membrane; c, liquid outlet; d, sculpted optic fiber. Middle, Probe tip detail. Scale bar, 0.2 mm. Right, Picture of the optogenetic-microdialysis probe showing light spread pattern at the probe tip. Scale bar, 1 mm. ***C***, ***D***, Time course of extracellular concentrations of (***C***) DA (red) and (***D***) ACh (black) in the NAc. ***E***, ***F***, Time course of extracellular concentrations of (***E***) DA (red) and (***F***) glutamate (Glu, green) in the NAc with constant perfusion (reverse dialysis) of the nAChR antagonist DHβE (10 μM). Time 0-60 represents the values of samples before stimulation. Blue vertical lines indicate the period of optogenetic stimulation (20 min). Results are expressed as mean ± SEM of percentage of the average of three values before stimulation. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

the nAChR antagonist DHβE (10 μM) delivered locally via reverse dialysis, the same stimulation protocol did not cause any change in extracellular DA concentration in the striatal dialysates (96.5 ± 11.6% of baseline, $n = 9$; $F = 0.64$, $p = 0.70$, repeated-measures one-way ANOVA; Fig. 1*E*), indicating that activation of nAChRs in the striatum is required for the cortically evoked DA release *in vivo*. As a control, we measured glutamate concentration in these same dialysates, which showed increased glutamate levels and confirmed the delivery of cortical stimulation in the presence of the nAChR antagonist (184.3 ± 27.3% of baseline, $n = 9$; $F_{(6,45)} = 4.01$, $p = 0.003$, repeated-measures one-way ANOVA with mixed-effects model; $p = 0.0006$ for 60 min vs 80 min; Dunnett's multiple comparisons test; Fig. 1*F*).

Stimulation of PrL/IL inputs to striatum is sufficient to evoke local DA signals in brain slices

In order to precisely access the specific inputs and apply pharmacological agents, we used transgenic mice with combination of optogenetic stimulation. DA signals were recorded in the striatum using FSCV in mouse brain slice preparations. Release of DA from axonal projections within the striatum was evoked by electrical stimulation and compared with DA signals evoked by

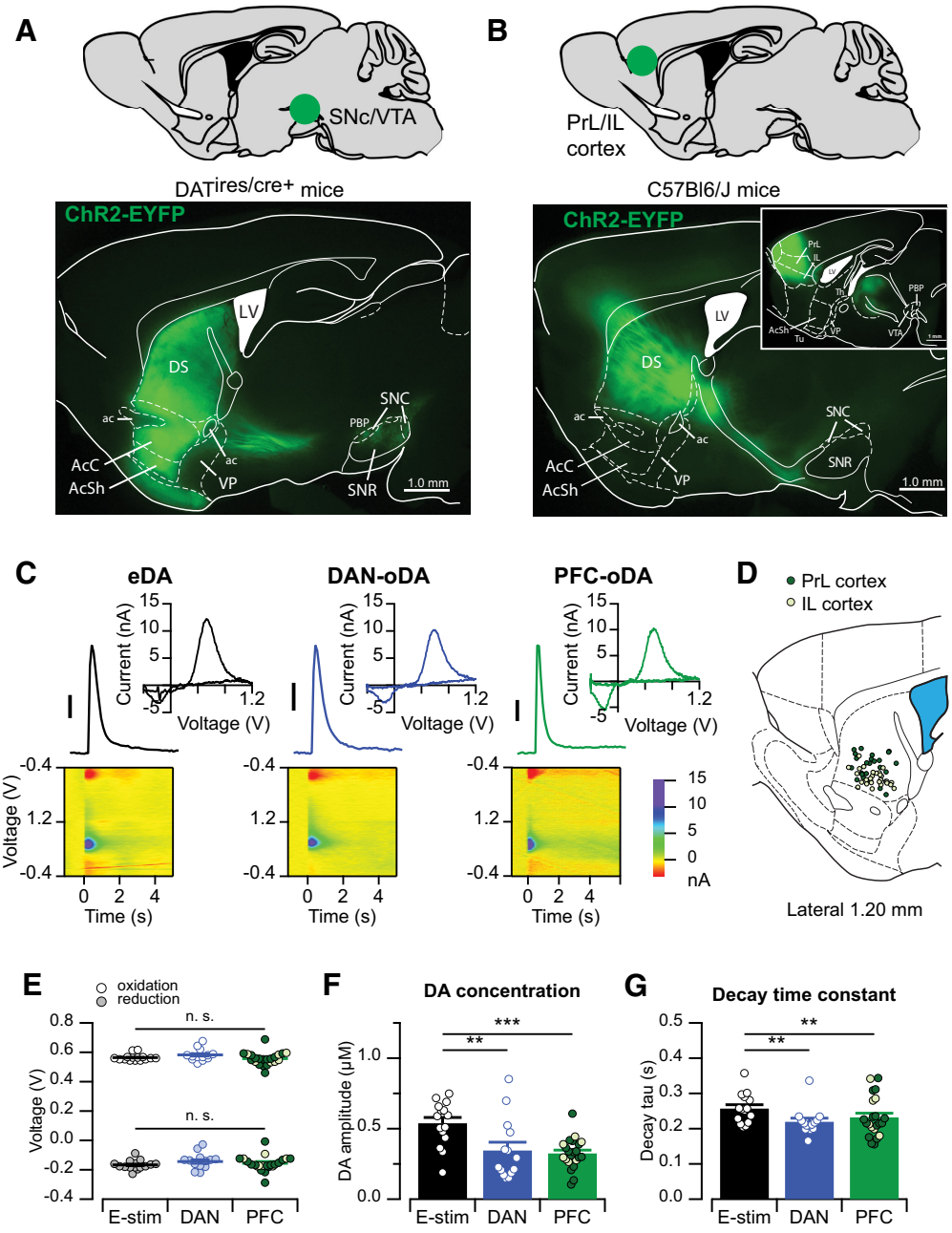


Figure 2. Striatal DA signals evoked by midbrain and cortical inputs. ***A***, Example of the fluorescence pattern (bottom) observed in a sagittal brain slice from a $DAT^{IRES\text{-}Cre+}$ mouse injected with DIO-ChR2-EYFP in the midbrain (top). ***B***, Example of the fluorescence pattern (bottom) observed in a sagittal brain slice from a C57Bl6/J ($DAT^{IRES\text{-}Cre-}$) mouse injected with ChR2-EYFP in the PrL or IL cortex (top). Inset, A more medial slice with the site of injection. ***C***, Representative DA transients, current–voltage (CV) plots, and color voltammograms when evoked by electrical stimulation (eDA, left), optogenetic stimulation of DAN fibers (DAN-oDA, middle), or optogenetic stimulation of PFC inputs (PFC-oDA, right). All three CV plots show the electrochemical profile of DA oxidation. Scale bar, 100 nM. ***D***, A sagittal brain section modified from the Mouse Brain Atlas (Franklin and Paxinos, 2007) showing the FSCV recording sites from mice injected with ChR2-EYFP in the PrL (filled) and IL (empty) cortex. ***E***, The oxidation (top) and reduction (bottom) voltages from the CV plots of eDA (left), DAN-oDA (middle), and PFC-oDA (right) were plotted as average with SEM. Open and filled circles represent individual values. ***F***, DA peak concentrations and ***G***, decay time constants of eDA (left), DAN-oDA (middle), and PFC-oDA (right) were plotted as average with SEM. Open circles represent individual values. For PFC-oDAs, the same color code was applied as in ***D*** for PrL and IL according to injection location. ***E–G***, Data points include experiments where electrical and optogenetic stimulation was delivered alternatingly in the same slice or in different slices (see Materials and Methods). $^{**}p < 0.01$, $^{***}p < 0.001$. LV, Lateral ventricle; DS, dorsal striatum; ac, anterior commissure; AcC, accumbens core; AcSh, accumbens shell; VP, ventral pallidum; PBP, parabrachial pigmented nucleus; SNR, substantia nigra reticulata; Th, thalamus; Tu, olfactory tubercle. n.s., not significant.

input-specific optogenetic stimulation. To achieve input-specific stimulation, ChR2-EYFP was expressed in either midbrain DANs and their axonal projections to the striatum using Cre-dependent viral vector injection in $DAT^{IRES\text{-}Cre+}$ mice (Fig. 2*A*) or ChR2-EYFP in cortical neurons of the PrL/IL cortex and their projections to the striatum using viral injection in Cre-negative littermate mice (Fig. 2*B*).

As expected, single brief pulses of electrical stimulation delivered via an electrode placed within the striatal tissue reliably evoked DA (electrically evoked DA transient [eDA]) signals in striatal brain slices (Fig. 2*C*). Also, direct optogenetic stimulation of midbrain DAN axonal projections within the striatum also evoked DA signals (referred here as DAN-oDA) on delivery of a single brief pulse of blue light (0.6 ms), in agreement with

previous published studies (Adrover et al., 2014; Melchior et al., 2015). Last, similar to the *in vivo* findings in Figure 1, selective optogenetic stimulation of axon projections from PrL/IL cortex into the striatum was also sufficient to trigger DA signals (referred as PFC-oDA) in the striatum of in slice preparation (Fig. 2*C*). These PFC-oDA signals were also reliably evoked by brief single pulses of blue light (0.6 ms) in the ventral portion of the dorsomedial striatum in sagittal brain slices (Fig. 2*D*). The viral expression around injection site was often large enough to include both PrL and IL regions of the cortex. For this reason, we combined data from the two regions. It is worth noting that these brain sections mainly contain the axon projections from PrL/IL neurons, but not their cell bodies, which are localized to more medial sagittal sections of the brain (Fig. 2*B*, inset).

The current–voltage plots of DA signals evoked via the three different types of stimulation (electrical, optogenetic stimulation of DANs, and optogenetic stimulation of PrL/IL) were indistinguishable and displayed current peaks at the expected oxidation (0.56 ± 0.01 V for eDA, $n = 15/n = 8$; 0.58 ± 0.01 V for DAN-oDA, $n = 14/n = 6$; 0.56 ± 0.01 V for PFC-oDA, $n = 22/n = 13$; $F = 0.49$, $p = 0.62$, repeated-measures one-way ANOVA with mixed-effects model; Fig. 2*E*) and reduction voltages (-0.17 ± 0.01 V for eDA, -0.14 ± 0.01 V for DAN-oDA, -0.16 ± 0.02 V for PFC-oDA; $F = 0.76$, $p = 0.49$, repeated-measures one-way ANOVA with mixed-effects model; Fig. 2*E*) for DA. The peak amplitude of eDA was significantly higher than DAN-oDA or PFC-oDA (556 ± 42 nM for eDA; 367 ± 58 nM for DAN-oDA; 325 ± 23 nM for PFC-oDA; $F_{(2,12)} = 13.430.76$, $p > 0.05$, repeated-measures one-way ANOVA with mixed-effects model; $p = 0.01$ for eDA vs DAN-oDA, $p = 0.0005$ for eDA vs PFC-oDA, and $p = 0.99$ for DAN-oDA vs PFC-oDA; Tukey's multiple comparisons test; Fig. 2*F*), and the decay time constants (tau) were similarly longer with eDA (0.25 ± 0.01 s for eDA, 0.22 ± 0.01 s for DAN-oDA, 0.23 ± 0.01 s for PFC-oDA; $F_{(2,12)} = 13.43$, $p = 0.0009$, repeated-measures one-way ANOVA with mixed-effects model; $p = 0.004$ for eDA vs DAN-oDA, $p = 0.009$ for eDA vs PFC-oDA, and $p = 0.83$ for DAN-oDA vs PFC-oDA; Tukey's multiple comparisons test; Fig. 2*G*). Thus, we were able to reliably evoke steady DA signals by selective optogenetic stimulation of axon projections from PrL/IL cortex in the ventral portion of the dorsomedial striatum. The DA signals recorded using FSCV showed indistinguishable chemical and temporal characteristics, even when evoked by different inputs.

Unique physical and pharmacological properties across input-specific DA signals

We first explored the threshold for triggering the input-specific optogenetically evoked DA signals. The threshold was determined by constructing input-output curves and varying the light pulse durations from 0.1 to 5 ms in brain slices from mice expressing ChR2 in midbrain DANs or the PrL/IL region (Fig. 3*A*,*B*). The relative amplitudes of DA transients were overlapping for pulses of 0.5-5 ms duration, and both DAN-oDA and PFC-oDA signals showed maximal amplitude with 1-5 ms light pulse durations. However, very short light pulses (0.1 ms) evoked measurable DA signals only in slices in which ChR2 was expressed in midbrain DAN projections to the striatum, indicating a lower threshold for evoking DA signals with direct stimulation of DA neuron axons than with stimulation of cortical inputs (duration × input, $F_{(5,90)} = 11.58$, $p < 0.0001$, repeated-measures two-way ANOVA; *post hoc* for the duration of 0.1 ms: $33 \pm 4\%$ for DAN-oDA, $n = 12/n = 8$, and $1.9 \pm 0.7\%$ for PFC-oDA, $n = 8/n = 5$, $p < 0.0001$; Fig. 3*B*). A higher threshold for the cortically evoked DA signals could reflect technical differences, such as opsin expression levels, or other biological factors, such as the density of innervation of cortical versus DAN axonal projections, difference in the release probability of the DA varicosities that respond to direct stimulation of DAN axons from those responding to cortical inputs, and/or the indirect and polysynaptic nature of the cortical-evoked DA signals.

To determine whether PrL/IL activation elicits striatal DA release and engages the CIN-dependent mechanism as shown in *in vivo* microdialysis experiment (Fig. 1), we then studied the basic pharmacological characteristic of DA signals evoked by stimulation of PFC inputs and compared them with those evoked by electrical and DAN input stimulation. In striatal brain slices, PFC-oDA signals were completely abolished by a mixture of AMPA and NMDA receptor antagonists (NBQX/CPP), whereas eDA and DAN-oDA signals were unaffected (amplitude after the antagonists = $0.3 \pm 1.6\%$ of baseline for PFC-oDA, $n = 6/n = 5$; $107.3 \pm 3.0\%$ for eDA, $n = 6/n = 4$; $99.8 \pm 1.4\%$ for DAN-oDA, $n = 6/n = 4$; $F_{(2,3)} = 694.8$, $p < 0.0001$, repeated-measures one-way ANOVA with mixed-effects model; $p = 0.14$ for eDA vs DAN-oDA, p values < 0.0001 for PFC-oDA vs eDA or PFC-oDA vs DAN-oDA; Tukey's multiple comparisons test; Fig. 3*C*, *D*). These findings indicate a requirement for ionotropic glutamate transmission for PrL/IL inputs to evoke local DA signals, which is in agreement with previous reports on the requirement of glutamate transmission for DA signals evoked by motor cortex and thalamic inputs to the striatum (Threlfell et al., 2012; Kosillo et al., 2016; Cover et al., 2019). Further, similar to our *in vivo* data (Fig. 1*D*), DHβE also abolished PFC-oDA signals, and significantly depressed eDA signals, while having no effect on DAN-oDA signals ($1.9 \pm 1.4\%$ of baseline for PFC-oDA, $n = 11/n = 7$; $21 \pm 3\%$ for eDA, $n = 10/n = 5$; $99 \pm 3\%$ for DAN-oDA, $n = 5/n = 2$; $F_{(2,7)} = 395.2$, $p < 0.0001$, repeated-measures one-way ANOVA with mixed-effects model; $p < 0.0001$ for eDA vs DAN-oDA, $p = 0.0006$ for eDA vs PFC-oDA, and $p < 0.0001$ for DAN-oDA vs PFC-oDA; Tukey's multiple comparisons test; Fig. 3*E*,*F*). Together, the pharmacology results support the hypothesis that PFC-oDA signals are mediated by activation of glutamate synapses from PrL/IL cortex to striatal CINs, which excites them to fire and trigger ACh release, thereby driving DA release via activation of nAChRs on DAN axon fibers within the striatum.

ACh is rapidly cleared by enzymatic degradation by AChE (Quinn, 1987). The extremely dense presence of AChE in the striatum (Zhou et al., 2001) assures the termination of cholinergic action. We also have previously shown the influence of the AChE on the DA transmission in the striatum (Shin et al., 2015, 2017). As AChE activity shows temperature dependence (Vidal et al., 1987), we then tested the temperature dependence of the input-specific DA transients by varying the temperature of the bath solution. The amplitude of both DAN-oDA and PFC-oDA signals increased linearly as the temperature was lowered from the standard 32°C to 25°C, and both showed a tight negative regression with similar slopes significantly different from zero ($-10.8 \pm 0.6\%$ per °C, $n = 8/n = 3$, for DAN-oDA, and $-10.8 \pm 2.0\%$ per °C, $n = 6/n = 3$, for PFC-oDA; p values < 0.0001; Fig. 3*G*,*H*). The temperature dependence of the DA signal amplitude likely reflects the change in rate of DA transporter activity with temperature. DA transporter activity was reported to have a temperature coefficient Q_{10} ranging from 1.39 to 2.95 (Zhu and Hexum, 1992), from which we estimate a Q_{10} of ~2.3 for the DA transporter for 24°C–37°C range. Our estimate predicts that large changes in the rate of DA reuptake by the transporter would be around the tested and physiological temperatures.

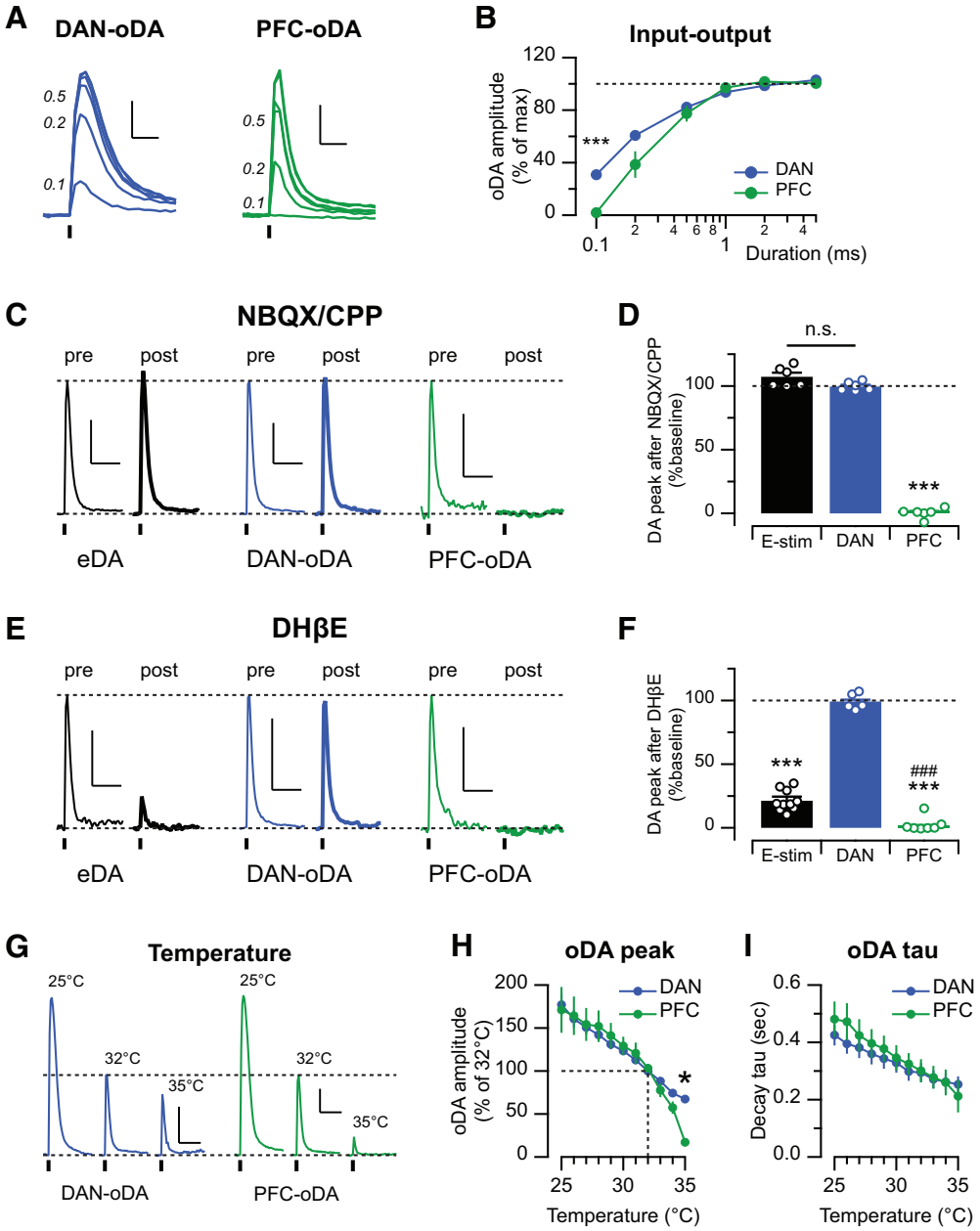


Figure 3. DAN-oDA and PFC-oDA show different physical and pharmacological properties. ***A***, Representative DAN-oDA (left) and PFC-oDA (right) transients evoked with different light pulse duration (in ms). Scale bar, 100 nM, 0.5 s. ***B***, oDA amplitudes normalized to their maximum response were averaged and plotted as a function of the stimulus duration. ***C***, Representative traces of eDA, DAN-oDA, and PFC-oDA transients before and after bath application of the glutamate receptor antagonists, NBQX and CPP (both 5 μM). Dotted line (top) indicates the amplitude of DA transients before the drugs. Scale bar, 200 nM, 2 s. ***D***, Averages with SEM of DA amplitude after NBQX and CPP were plotted. Open circles represent individual value. ***E***, Representative traces of eDA, DAN-oDA, and PFC-oDA transients before and after bath application of the β2-contatining nAChR antagonist, DHβE (1 μM). Scale bar, 200 nM, 2 s. ***F***, Averages with SEM of DA amplitude after DHβE were plotted. Open circles represent individual value. ***Versus DAN-oDA. $^{\#\#\#}$Versus eDA. ***G***, Representative traces of DAN-oDA (left) and PFC-oDA (right) transients at 25°C, 32°C, and 35°C. Dotted line (top) indicates the oDA amplitude at 32°C. Scale bar, 200 nM, 2 s. ***H***, Average oDA peak amplitudes normalized to 32°C. ***I***, Average oDA decay time constants were plotted as a function of temperature. $^{*}p < 0.05$, $^{***}p < 0.001$, $^{\#\#\#}p < 0.001$, n.s., not significant.

Indeed, the decay of the DA transients, which is a parameter that reflects the clearance of extracellular DA by the transporter, was also affected as the temperature was lowered. The decay time constant of both DAN-oDA and PFC-oDA increased linearly with the decrease of the temperature (DAN-oDA: -0.017 ± 0.001 s/°C; PFC-oDA: -0.025 ± 0.001 s/°C; Fig. 3*I*), which corresponded with the linear increase in amplitude. However, raising the temperature from 33°C to 35°C caused a dramatic and selective drop in the amplitude of the PFC-oDA signals, compared with DAN-oDA signals (Fig. 3*G*,*H*). The slope for PFC-oDA signals increased threefold from −11% to −31% per °C.

Vidal et al. (1987) have reported on a high dependence with temperature of the activity of AChE from rat brain, which reaches maximal rate at ~35°C. Our *in vivo* and *in vitro* pharmacology experiments showed that PFC-oDA signals require activation of nAChR. Therefore, we speculate that the observed steep temperature dependence of the signals is in large part mediated by rate increase in AChE activity. Furthermore, at physiological temperatures, the PFC-evoked DA signals may be very localized near ACh release sites, presumably overlapping with nAChR expression in DAN axons.

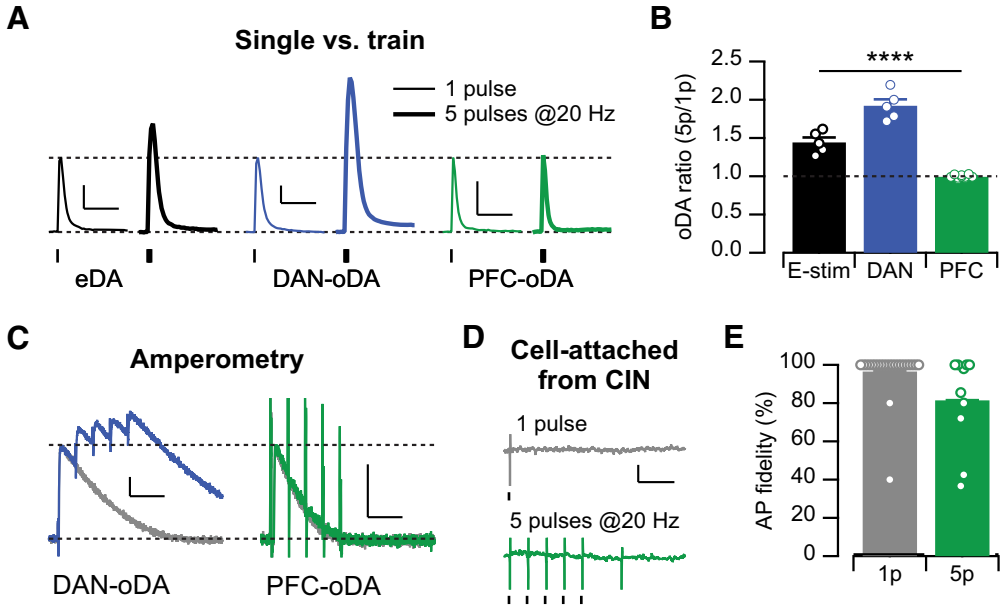


Figure 4. PFC-oDA shows no summation by train stimulations. ***A***, Representative DA traces of eDA, DAN-oDA, and PFC-oDA transients evoked by single pulse (1p, thin traces) or train of 5 pulses at 20 Hz (5p, thick traces). Scale bar, 200 nM, 2 s. ***B***, Averages with SEM of the DA amplitude ratio (5p/1p) for eDA, DAN-oDA, and PFC-oDA transients were plotted. Open circles represent individual values. ***C***, Representative amperometric traces for DAN-oDA (left) and PFC-oDA (right) transients evoked by single pulse (gray traces) or train of 5 pulses at 20 Hz (color traces). PFC-oDA amperometric transients evoked by 50 at 20 Hz were indistinguishable from 1p stimulation, except the large deflection at the stimulation time for 5p pulses. Scale bars: 200 pA, 100 ms. ***D***, Representative cell-attached recordings from CINs with single (top, gray) or train of 5 pulses at 20 Hz (bottom, green). Scale bars: 20 pA, 100 ms. ***E***, Averages with SEM of the action potential fidelity were plotted for the single pulse and train stimulation. Open circles represent individual values. ****$p < 0.0001$.

Activity dependent summation of the midbrain and cortical DA signals

Trains of stimulation pulses have been shown to produce sublinear summation in the amplitude of DA signals evoked by electrical and optogenetic stimulation of DAN axon projections in the striatum *in vitro* (Zhang and Sulzer, 2004; Threlfell et al., 2010; Melchior et al., 2015; Shin et al., 2017). We then tested the degree of summation of the DA signals in response to short trains of stimulation pulses (5 pulses at 20 Hz) and compared it with the transients evoked by a single pulse (Fig. 4*A*,*B*). This pattern of train stimulation was chosen based on burst firing patterns recorded *in vivo* for DANs during behavior (Schultz et al., 1993; Hyland et al., 2002). Indeed, we found that trains of 5 pulses at 20 Hz evoke DA transients 50% larger than those evoked by single pulse stimulation for eDA signals and almost double for DAN-oDA (5p/1p: 1.44 ± 0.06, $n = 5/n = 3$, for eDA; 1.92 ± 0.19, $n = 5/n = 2$, for DAN-oDA; 1.00 ± 0.02, $n = 10/n = 7$; $F_{(2,2)} = 110.6$, $p = 0.009$, repeated-measures one-way ANOVA with mixed-effects model; $p = 0.04$ for eDA vs DAN-oDA, $p = 0.03$ for eDA vs PFC-oDA, and $p = 0.008$ for DAN-oDA vs PFC-oDA; Tukey's multiple comparisons test; Fig. 4*A*,*B*). On the contrary, PFC-oDA transients evoked by this train stimulation had similar amplitudes compared with those evoked by a single pulse, indicating no summation of DA signals evoked by PrL/IL inputs. In order to measure DA signals with better temporal resolution, we performed amperometric recordings and again compared DA signals evoked by single pulse and trains (Fig. 4*C*). In line with the FSCV results, amperometric recordings showed that DAN-oDA signals display sublinear summation, whereas PFC-oDA signals are indistinguishable between single pulse or trains of 5 pulses at 20 Hz. The lack of summation in the PFC-oDA signals in response to trains resembles the findings obtained with DA signals evoked by synchronized activation of CINs (Threlfell et al., 2012; Shin et al., 2017) and further supports the idea that the PFC-evoked local DA signals are mediated by synchronized action of CINs in response to stimulation of PrL/IL inputs. The large transient current deflection in response to each pulse of stimulation ($n = 5/n = 4$; Fig. 4*C*, right) is suggestive of the action potential firing evoked in CINs and other striatal neurons by the optogenetic stimulation of PrL/IL inputs. The fact that every pulse of the train stimulation triggers a current deflection reflects the ability of PrL/IL inputs to evoke action potentials firing in downstream neurons and supports the idea that the lack of summation in DA signals is not likely because of the failure of action potential firing by CINs. To directly test the contribution of failure of action potential firing by CINs to the lack of summation in PFC-oDA signals, we conducted cell-attached recordings from CINs and measured action potential fidelity in response to a single pulse and trains of 5 pulses at 20 Hz (single pulse: $96.5 \pm 2.7\%$, $n = 23/n = 4$, $t = 1.28$, $p = 0.21$; trains of 5: $81.5 \pm 7.7\%$, $n = 10/n = 4$, $t = 2.42$, $p = 0.04$; both with one-sample t test to 100% value; Fig. 4*D*,*E*). Although there was a small but significant percentage of failure, the totality of the data indicates that the lack of summation is not likely caused by action potential failure, but instead can be attributed to other downstream mechanisms, such as desensitization of nAChRs (Threlfell et al., 2012; Shin et al., 2017).

Input-specific DA signals are evoked by dual optogenetic stimulation in the same brain slice

Thus far, DA signals evoked by the different inputs were recorded in brain slices from different mice that expressed ChR2 in either midbrain DANs or in PrL/IL cortex. In order to strictly test the segregation/distinction of these two pathways, it was important to examine these two inputs and compare the input-specific DA signals in the slice preparation from the same mice. We took advantage of opsins activated by different light wavelengths, specifically, ChR2 and the red-shifted opsin ChrimsonR

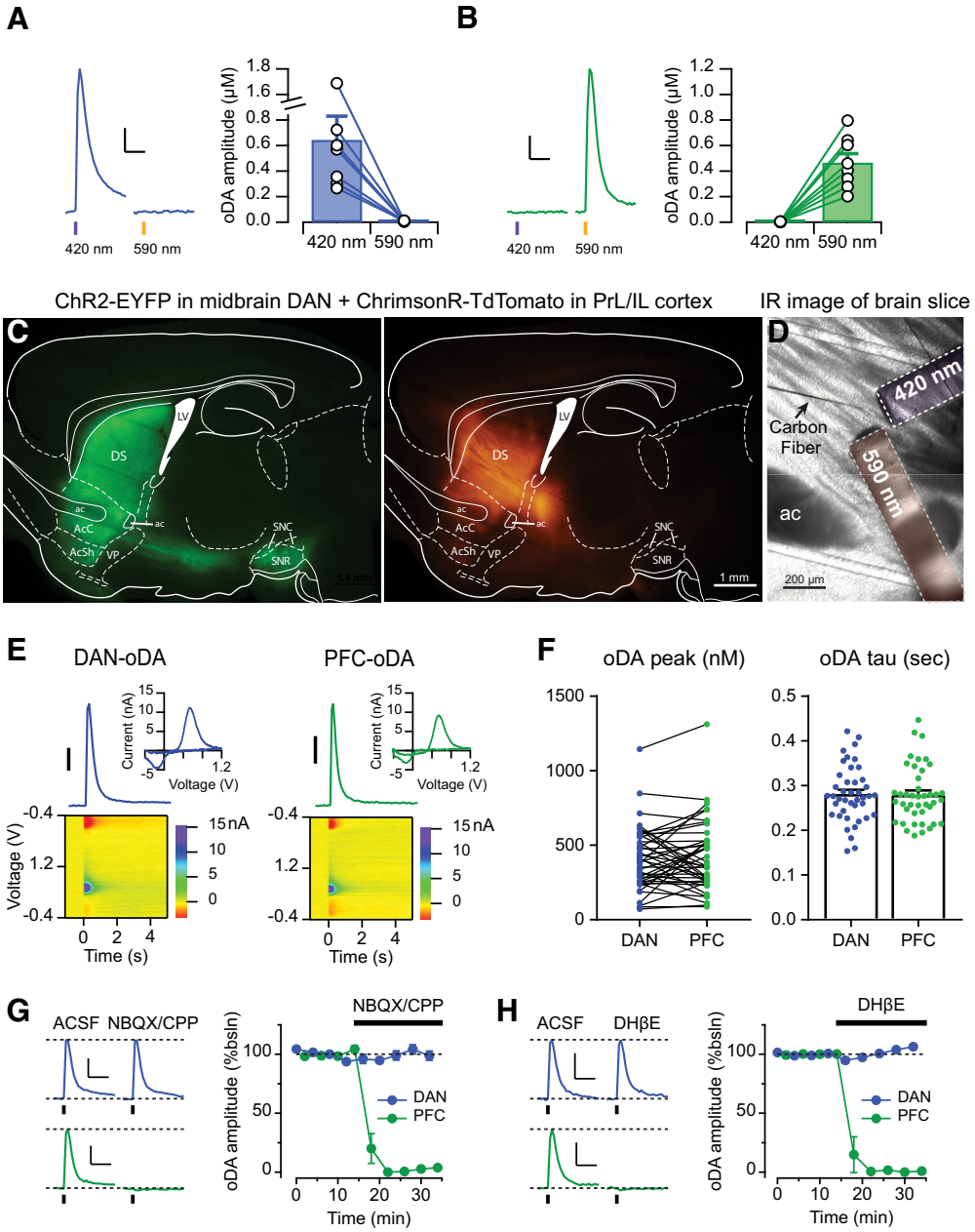


Figure 5. Dual-opsin expression to evoke cortical and midbrain DA signals in the same brain slices. ***A***, ***B***, Left, Representative traces of DA signals evoked by either 420 or 590 nm light pulses from mice expressing (***A***) only ChR2 in midbrain DANs or (***B***) only ChrimsonR in PrL/IL cortex. Right, oDA amplitudes were plotted as pairs for each wavelength. Scale bars: 200 nM for A 100 nM for ***B***; 1s ***C***, Example of the fluorescence patterns with filter set for yellow signal (left) or red signal (right) from a sagittal brain slice of a $DAT^{IRES\text{-}Cre+}$ mice injected with ChrimsonR-TdTomato in the PFC and DIO-ChR2-EYFP in the midbrain. ***D***, Configuration to the FSCV DA recording using a carbon fiber and two fiber-optics delivering 420 and 590 nm, respectively. ***E***, Representative DA transients, CV plots, and color voltammograms of DAN-oDA and PFC-oDA. Scale bar: 100 nM ***F***, Left, Amplitudes were plotted as pairs for DAN-oDA and PFC-oDA recorded from the same slices. Right, Averages with SEM of decay time constant were plotted for DAN-oDA and PFC-oDA. Dots represent individual values. ***G***, Left, Representative traces of DAN-oDA (top) and PFC-oDA (bottom) before and after the application of glutamate receptor antagonists NBQX and CPP. Right, Averages with SEM of DAN-oDA and PFC-oDA were plotted as a function of time as NBQX/CPP was applied. ***H***, Left, Representative traces of DAN-oDA (top) and PFC-oDA (bottom) before and after the application of nAChR antagonists DHβE. Right, Averages with SEM of DAN-oDA and PFC-oDA were plotted as a function of time as DHβE was applied. Scale bars for ***G***–***H***: 100 nM; 1 s.

(Klapoetke et al., 2014). We first set up the conditions for selective stimulation of each opsin without cross-activation. In brain slices expressing only ChR2 in midbrain DANs, pulses of purple light (420 nm) evoked reliable DA signal, whereas similar and longer pulses of orange light (590 nm) did not trigger any detectable signals (Fig. 5*A*). Conversely, in slices expressing only ChrimsonR in the PrL/IL cortex, brief orange light pulses evoked DA signal, but not purple light (Fig. 5*B*). Thus, under these experimental conditions, ChR2 and ChrimsonR can be used in combination to selectively stimulate two different inputs by delivering pulses of two different light wavelengths without any detectable cross-activation.

Next, in the same $DAT^{IRES\text{-}Cre+}$ mice, ChR2-YFP was expressed in midbrain DANs with a Cre-dependent expression vector, and ChrimsonR-TdTomato was expressed in PrL/IL cortex. Figure 5*C* shows examples of the fluorescent expression pattern of ChR2-YFP (green) and ChrimsonR (red) in the same sagittal brain slice from this double-injected mouse. Figure 5*D* shows the experimental arrangement for the dual-input recordings with the placement of the carbon fiber and the two fiber-

optics for delivering the purple and orange light pulses that were used to activate ChR2 and ChrimsonR, respectively. In these slices with dual-opsin expression, brief pulses of either purple or orange light evoked DA signals with indistinguishable current–voltage plots, showing the characteristic peaks for DA oxidation and reduction (Fig. 5*E*). The DA concentration transients had comparable peak amplitudes with means of 407 ± 32 nM for DAN-oDA and 420 ± 36 nM for PFC-oDA ($n = 43/n = 13$, $t = 0.47$, df = 42, $p = 0.64$, paired t test; Fig. 5*F*) and overlapping time courses as measured at 10 Hz sample rate of FSCV, in agreement with the results shown in Figure 2. The decay time constant of the DA transients was similar (0.28 ± 0.01 s for DAN-oDA and 0.28 ± 0.01 s for PFC-oDA, $n = 43/n = 13$, $t = 0.41$, df = 41, $p = 0.68$ from paired t test; Fig. 5*F*).

Then, we tested the pharmacological properties of DAN-oDA and PFC-oDA by delivering alternating purple and orange light pulses and recording from a single carbon fiber. Application of the glutamate receptor antagonists NBQX/CPP (99.2 ± 3.1% of the baseline for DAN-oDA and 1.6 ± 1.1% for PFC-oDA, $n = 5/n = 3$; $t = 25.3$, df = 4, $p < 0.0001$, two-tailed paired t test; Fig. 5*G*) or the nAChR antagonist DHβE (101.9 ± 0.6% of the baseline for DAN-oDA and 0.8 ± 0.8% for PFC-oDA, $n = 5/n = 3$; $t = 79.13$, df = 4, $p < 0.0001$, two-tailed paired t test; Fig. 5*H*) completely abolished PFC-oDA signals while leaving DAN-oDA signals in the same location intact. These results confirm the findings from Figure 3, which use antagonists and single-input optogenetic stimulation to evoke either DAN-oDA or PFC-oDA in separate slices/mice. These results also validate the experimental approach by showing input specificity with no apparent crossover in the optogenetic stimulation of ChR2 and ChrimsonR with the described wavelengths. More importantly, these pharmacological findings highlight the different mechanisms and the unique nature of the input-specific DA signals evoked by PFC and midbrain inputs.

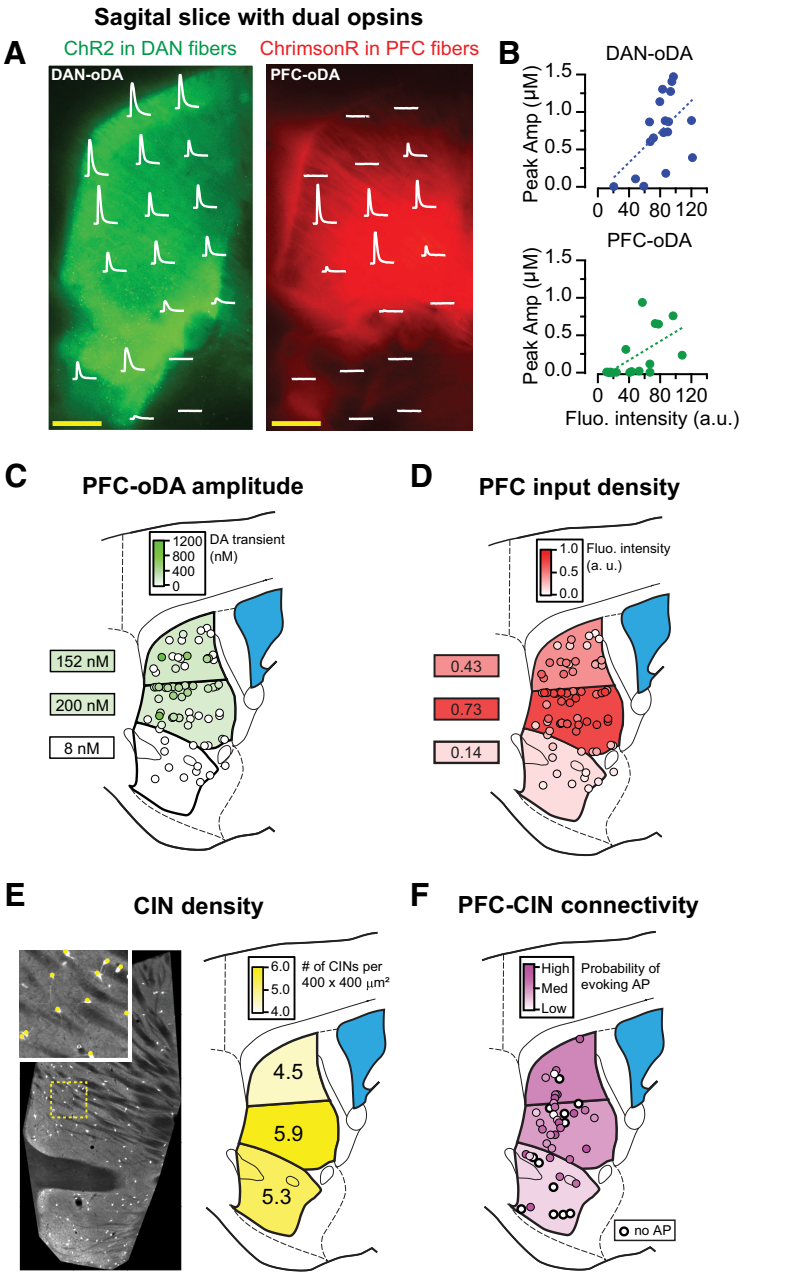


Figure 6. Cortically evoked DA signals are spatially restricted. ***A***, The locations of carbon fiber and two fiber-optics were adjusted to measure DAN-oDA (left) and PFC-oDA (right) from each location where the corresponding oDAs were superimposed on the fluorescence patterns. Scale bars, 500 μm. ***B***, Peak amplitudes of DAN-oDA (top) and PFC-oDA (bottom) were plotted as a function of florescence intensity at each location shown in ***A***. Dotted lines indicate linear regression between the oDA amplitudes and the fluorescence intensities. ***C***, PFC-oDAs were measured from 74 locations (between ~1.0 and 1.2 mm in mediolateral coordinate) of mice injected with ChR2-EYFP in PrL/IL cortex and color-coded according to their peak DA concentrations. The average PFC-oDA amplitudes were calculated and color-coded for the three subregions. ***D***, The PFC input fluorescence intensities were color-coded for the same 74 locations in ***C***. The averages from the three subregions were calculated and color-coded. ***E***, Left, Fluorescence image of striatal CINs labeled with td-Tomato. Inset, Examples of identified CINs (red) from the area with the dotted yellow line. Right, The average numbers of CINs per 400 × 400 μm^2 (area as shown in the inset) were calculated and shown for the three subregions. ***F***, Using cell-attached patch recording, action potentials evoked by PFC stimulation were observed from a total of 46 CINs. For each CIN, a minimum light intensity to evoke action potentials was determined to score the connectivity. White circles with a thicker line represent CINs which did not show any evoked action potentials even with the maximum light intensity.

Cortically evoked signals are spatially restricted

As shown in Figure 2*D*, DA signals evoked by selective stimulation of PrL/IL inputs were preferentially detected between the NAc core and dorsal striatum. We then set out to investigate in more detail the spatial distribution of the input-specific DA signals using the dual-opsin expression approach by sampling the amplitude of DA signals across 18 different striatal areas of the sagittal brain slice while evoking responses with purple and orange light. While DAN-oDA signals were detected throughout the whole striatum, PFC-oDA signals were spatially restricted (Fig. 6*A*,*B*). Both DAN-oDA and PFC-oDA signals showed significant correlations with the fluorescent intensity of the projections ($p = 0.03$ and $r^2 = 0.28$ for DAN-oDA; $p = 0.01$, $r^2 = 0.37$ for PFC-oDA; Fig. 6*B*). However, the slope of the linear regression was larger for DAN-oDA signals than PFC-oDA signals (10 vs 6

nM/AU, respectively; Fig. 6*B*). Thus, the intensity of the DAN-labeled fibers was a good predictor of the magnitude of DAN-oDA signals. On the other hand, the presence of cortical axon fluorescence was not always a predictor of PFC-oDA signals, likely because of the larger density of passing cortical axons that are not synaptic terminal run through the striatum.

Averaging the area responses in 5 brain slices from 3 mice, we found that the mean amplitude of the PFC-oDA signals was largest in the ventral part of the dorsal striatum, which again maps in gross terms with the location of the brightest intensity of ChR2-EYFP-labeled projections from PrL/IL cortex (Fig. 6*C*,*D*). However, the correlation between fluorescence intensity and amplitude of PFC-oDA signals is not always strong; for example, the more intensely labeled caudal areas near the globus pallidus show almost no PFC-oDA signal (Fig. 6*C*,*D*). The evidence of small or no PFC-oDA signals in the more caudal striatum could possibly reflect greater fiber density rather than innervation of the caudal portion of the striatum by the axonal projections from PrL/IL cortex.

The number of CINs is <1% of the total striatal neurons, and they are sparsely distributed throughout the striatum (Burke et al., 2017). We then quantified the density of CINs from the same area to further determine whether the distribution of CINs contribute to the spatial restriction of PFC-oDA signals. For this purpose, we took advantage of CIN-tdTomato mouse line, which fluorescently labels CINs throughout the striatal slices (Fig. 6*E*). The quantification analysis of cell numbers in 400×400 μm^2 showed that the density of CINs was also highest in the middle part of the dorsomedial striatum (6 slices from 3 mice), suggesting that an uneven distribution of CINs could also contribute to the spatial restriction of the local DA signals evoked by PrL/IL cortical inputs to the ventral portion of the dorsomedial striatum. To further test whether the spatial profile of PFC-oDA signals is in part determined by the strength of the innervation of CINs by PrL/IL cortical neurons, we performed cell-attached recordings from the same CINs of CIN-TdTomato mice and measured the ability of PFC optogenetic stimulation to evoke action potentials in CINs throughout the striatum (46 CINs from 3 mice). We found that the connectivity followed a similar pattern to the PFC-oDA profile, and the probability that CINs will fire an action potential in response to PFC stimulation was higher in the central part of the striatum than in the NAc and dorsal regions of the dorsal striatum (Fig. 6*F*).

In conclusion, this study offers a series of *in vitro* and *in vivo* evidence that stimulation of cortical inputs from the PrL/IL cortex can evoke DA release in the striatum via a local mechanism that recruits CINs and requires activation of nAChRs. Since these cortically evoked DA signals are spatially restricted and have different properties, we speculate that they are engaged in distinctive striatal functions from those assigned for DA signals evoked by midbrain DANs.

Discussion

This study provides strong evidence that DA signals in the striatum can occur both *in vivo* and *in vitro* in response to stimulation of PrL/IL cortical inputs. These PrL/IL-evoked DA signals have distinctive pharmacological and physiological properties, compared with the DA signals evoked by DAN inputs. The *in vitro* experiments also introduce a dual optogenetic stimulation approach with which we can activate two different inputs to the striatum in the same brain slice, without apparent cross-activation. This study further reveals the spatial localization of the DA signals evoked by PFC inputs.

We showed that the PrL/IL-evoked DA signals require activation of glutamate and nicotinic ACh receptors, in line of a series of elegant published work. Using *in vitro* radioactive assays and *in vivo* microdialysis, it was first shown that locally administered ACh and glutamate can trigger DA release in the striatum (Giorguieff et al., 1976, 1977). Giorguieff et al. (1976) showed that an nAChR antagonist blocked ACh-evoked DA signals, and speculated that ". . . the release of DA from dopaminergic terminals can be regulated by cholinergic presynaptic receptors exhibiting nicotinic characteristics." To their credit, their conclusions agree with our interpretations of the findings from the current study as well as other recent optogenetic studies where it was shown that selective stimulation of CINs is sufficient to evoke DA signals (Cachope et al., 2012; Threlfell et al., 2012; Wang et al., 2014; Shin et al., 2017). The involvement of CINs in the glutamate-dependent DA signals was also suggested by early work by Taber and Fibiger (1994) who showed that electrical stimulation of PFC can increase levels of ACh in the striatum and further supported by more recent optogenetic studies (Threlfell et al., 2012; Kosillo et al., 2016; Johnson et al., 2017; Mateo et al., 2017; Cover et al., 2019).

Here, our experiments confirmed the requirement for ionotropic glutamate receptor activation in the PFC-evoked DA signals as previously shown (Mateo et al., 2017). We also show that nAChRs are required for the PrL/IL-evoked DA signals *in vitro*, similar to other findings when stimulating inputs from other cortical (Kosillo et al., 2016) and thalamic areas (Threlfell et al., 2012; Kosillo et al., 2016; Cover et al., 2019) and indirectly *in vivo* (Mateo et al., 2017). Our results from the *in vivo* microdialysis experiments reproduce the original findings from Quiroz et al. (2016) showing that optogenetic stimulation of PFC inputs increases striatal levels of DA *in vivo* (and glutamate, as expected; Fig. 1). More importantly, we showed that stimulation of PrL/IL inputs to the striatum also elevates striatal levels of ACh *in vivo* and that this cortically evoked DA signals are blocked when a nAChR antagonist is perfused (Fig. 1*C*,*D*). Thus, these findings support that idea PrL/IL inputs form excitatory synapses on striatal CINs and can activate then to evoke ACh release (Fig. 1*C*). In agreement with this conclusion, recent work showed that optogenetic stimulation of either M1 motor cortex or parafascicular nucleus of the thalamus can induce release of ACh that was detected using exogenous G-protein-coupled inward rectifying K-current expressed in medium spiny neurons (Mamaligas et al., 2019). Together, our findings from both *in vitro* and *in vivo* strongly support the hypothesis of a local mechanism for evoking DA release in the striatum, in addition to the more conventional mechanism based on midbrain DAN firing. This local mechanism engages striatal CINs, the targets of inputs from PFC, which have an essential role in evoking DA release from DAN fibers.

DA signals evoked by PrL/IL inputs and those evoked by DAN fibers stimulation share common properties, such as overlapping electrochemical profiles of the voltammetric currents and similar concentration range, time course, and decay time constant of the signals (Figs. 2, 5*E*,*F*). However, there are also several differences between the input-specific DA signals. First, DAN-evoked DA signals do not require activation of either ionotropic glutamate receptors or nAChRs (Figs. 3*C*–*F*, 5*G*,*H*), confirming the direct nature of this mechanism that is triggered by ChR2-evoked action potentials in DAN axon fibers. Second, PrL/IL-evoked DA signals display different temperature

sensitivity and a higher threshold compared with DAN-oDA, requiring a longer duration of light stimulation as determined in the input-output relationship (Figs. 3, 4). This finding may reflect the polysynaptic nature of the cortically evoked signals and the requirement of synchronized activation of CINs to trigger DA release by this local mechanism (Threlfell et al., 2012; Kosillo et al., 2016; Liu et al., 2018). Third, the amplitude of DAN-evoked DA signals increases as the number of stimulation pulses increases, indicating some summation in the DA signals evoked by a train of stimulation pulses (Fig. 4*A*,*B*). In contrast, the amplitude of DA signals in response to a train of pulses of cortical input stimulation is similar to the amplitudes obtained in response to a single pulse of stimulation, indicating no summation under this condition for the local mechanism (Fig. 4*A*,*B*). ChR2 displays use-dependent inactivation when stimulated at high frequency (Hass and Glickfeld, 2016). However, the lack of summation by the train stimulation is unlikely because of this use-dependent inactivation of ChR2 since 20 Hz train stimulation was still able to evoke action potentials in CINs recorded in cell-attached mode (Fig. 4*D*,*E*) and amperometry recording showed large deflections in response to train stimulation (Fig. 4*C*). Anecdotally, these current deflections disappeared when the ionotropic glutamate receptor antagonists were applied, also supporting the idea that firing in CINs and other downstream neurons occurs in response to 20 Hz train stimulation. The lack of summation was also reported in DA signals evoked by direct optogenetic stimulation of CINs in the dorsal striatum (Threlfell et al., 2012; Shin et al., 2017), which was suggested to be because of nAChR desensitization. We speculate then that the lack of summation we report here by 20 Hz trains in PrL/IL-evoked DA signals, which also require nAChRs activation, is also because of desensitization.

To directly study and compare the input-specific DA signals in the same animal, we established dual optogenetic stimulation using a red-shifted opsin, ChrimsonR, in combination with the blue-light activated ChR2. Two wavelengths for excitation were chosen based on the excitation spectrum of each opsin (Klapoetke et al., 2014), 590 nm for ChrimsonR and 420 nm for ChR2, which showed no cross-activation (Fig. 5*A*,*B*). This dual optogenetic approach was further validated using pharmacology in animals expressing ChR2 in midbrain DANs and ChrimsonR in PrL/IL neurons. We found that PrL/IL-evoked DA signals were completely blocked by antagonists of either ionotropic glutamate receptors or nAChRs, without affecting DAN-oDA signals in the same slice (Fig. 5*G*,*H*). Therefore, we determined that the two inputs can be activated independently using this dual optogenetic approach in brain slices from the same animal.

Interestingly, using this dual optogenetic approach, an intriguing spatial pattern was revealed for the cortically evoked PrL/IL DA signals. While DAN-oDA signals were recorded throughout the striatum, PrL/IL oDA signals were spatially restricted to a "hot-spot" region around the boundary between the NAc core and dorsal medial striatum (Fig. 6). This hot-spot area for the PrL/IL oDA signals corresponds to the same area where we found the highest density of CINs in our analysis. Thus, CIN density could contribute in part to the generation of this hot-spot area for the PrL/IL-evoked DA signals. Further, we also found that this area has the highest intensity of fluorescently labeled PrL/IL fibers, as well as the highest degree of synaptic connections to CINs, which is also in agreement with reported patterns of cortical connectivity (Voorn et al., 2004). Our initial interest was to study the PFC projections to the accumbens. However, since PrL/IL-evoked DA signals were largest and more reliable in the ventral DS (Fig. 6), recordings were performed in this area. Together, these findings suggest that PrL/IL inputs and the synaptic innervation to CINs are preferentially localized to the dorsomedial subregion, giving rise to the largest cortically evoked DA signals.

From the mechanism described here for this PrL/IL-evoked DA signals, we propose that any input to the striatum that strongly activates CINs (cortical, thalamic, or any other brain region) can in theory evoke local DA signals in the striatum. We also speculate, based on ours and other groups' works, that these glutamate-driven DA signals can similarly happen in other striatal subregions. The selective localization of PrL/IL-evoked DA signals suggests the existence of a "topographic map" for the local DA signals where inputs from different cortical subregions can trigger local DA signals at the striatal subregion which they innervate. Spatial segregation of input-specific DA signals in the striatum is a new concept introduced here that is rarely considered in the field. This concept may become helpful in understanding how DA signals in the striatum can be involved in so many diverse processes ranging from reinforcement learning and motivation to motor output and action selection.

References

Adrover MF, Shin JH, Alvarez VA (2014) Glutamate and dopamine transmission from midbrain dopamine neurons share similar release properties but are differentially affected by cocaine. J Neurosci 34:3183–3192.

Backman CM, Malik N, Zhang Y, Shan L, Grinberg A, Hoffer BJ, Westphal H, Tomac AC (2006) Characterization of a mouse strain expressing Cre recombinase from the 3′ untranslated region of the dopamine transporter locus. Genesis 44:383–390.

Burke DA, Rotstein HG, Alvarez VA (2017) Striatal local circuitry: a new framework for lateral inhibition. Neuron 96:267–284.

Cachope R, Mateo Y, Mathur BN, Irving J, Wang HL, Morales M, Lovinger DM, Cheer JF (2012) Selective activation of cholinergic interneurons enhances accumbal phasic dopamine release: setting the tone for reward processing. Cell Rep 2:33–41.

Chuhma N, Mingote S, Moore H, Rayport S (2014) Dopamine neurons control striatal cholinergic neurons via regionally heterogeneous dopamine and glutamate signaling. Neuron 81:901–912.

Cover KK, Gyawali U, Kerkhoff WG, Patton MH, Mu C, White MG, Marquardt AE, Roberts BM, Cheer JF, Mathur BN (2019) Activation of the rostral intralaminar thalamus drives reinforcement through striatal dopamine release. Cell Rep 26:1389–1398.e3.

Floresco SB, Yang CR, Phillips AG, Blaha CD (1998) Basolateral amygdala stimulation evokes glutamate receptor-dependent dopamine efflux in the nucleus accumbens of the anaesthetized rat. Eur J Neurosci 10:1241–1251.

Franklin KB, Paxinos G (2007) The mouse brain in stereotaxic coordinates, Ed 3. Amsterdam: Elsevier.

Gerfen CR (2000) Dopamine-mediated gene regulation in models of Parkinson's disease. Ann Neurol 47:S42–S50; discussion S50-S52.

Gerfen CR, Surmeier DJ (2011) Modulation of striatal projection systems by dopamine. Annu Rev Neurosci 34:441–466.

Giorguieff MF, Le Floc'h ML, Westfall TC, Glowinski J, Besson MJ (1976) Nicotinic effect of acetylcholine on the release of newly synthesized (3H) dopamine in rat striatal slices and cat caudate nucleus. Brain Res 106:117–131.

Giorguieff MF, Kemel ML, Glowinski J (1977) Presynaptic effect of L-glutamic acid on the release of dopamine in rat striatal slices. Neurosci Lett 6:73–77.

Hass CA, Glickfeld LL (2016) High-fidelity optical excitation of cortico-cortical projections at physiological frequencies. J Neurophysiol 116:2056–2066.

Hill DF, Parent KL, Atcherley CW, Cowen SL, Heien ML (2018) Differential release of dopamine in the nucleus accumbens evoked by low-versus high-frequency medial prefrontal cortex stimulation. Brain Stimul 11:426–434.

Hyland BI, Reynolds JN, Hay J, Perk CG, Miller R (2002) Firing modes of midbrain dopamine cells in the freely moving rat. Neuroscience 114:475–492.

Johnson KA, Mateo Y, Lovinger DM (2017) Metabotropic glutamate receptor 2 inhibits thalamically-driven glutamate and dopamine release in the dorsal striatum. Neuropharmacology 117:114–123.

Klapoetke NC, Murata Y, Kim SS, Pulver SR, Birdsey-Benson A, Cho YK, Morimoto TK, Chuong AS, Carpenter EJ, Tian Z, Wang J, Xie Y, Yan Z, Zhang Y, Chow BY, Surek B, Melkonian M, Jayaraman V, Constantine-Paton M, Wong GK, et al. (2014) Independent optical excitation of distinct neural populations. Nat Methods 11:338–346.

Kosillo P, Zhang YF, Threlfell S, Cragg SJ (2016) Cortical control of striatal dopamine transmission via striatal cholinergic interneurons. Cereb Cortex 26:4160–4169.

Lahiri AK, Bevan MD (2020) Dopaminergic transmission rapidly and persistently enhances excitability of D1 receptor-expressing striatal projection neurons. Neuron 106:277–290.e6.

Leviel V, Gobert A, Guibert B (1990) The glutamate-mediated release of dopamine in the rat striatum: further characterization of the dual excitatory-inhibitory function. Neuroscience 39:305–312.

Liu C, Kaeser PS (2019) Mechanisms and regulation of dopamine release. Curr Opin Neurobiol 57:46–53.

Liu C, Kershberg L, Wang J, Schneeberger S, Kaeser PS (2018) Dopamine secretion is mediated by sparse active zone-like release sites. Cell 172:706–718.e15.

Luscher C, Robbins TW, Everitt BJ (2020) The transition to compulsion in addiction. Nat Rev Neurosci 21:247–263.

Madisen L, Zwingman TA, Sunkin SM, Oh SW, Zariwala HA, Gu H, Ng LL, Palmiter RD, Hawrylycz MJ, Jones AR, Lein ES, Zeng H (2010) A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat Neurosci 13:133–140.

Mamaligas AA, Cai Y, Ford CP (2016) Nicotinic and opioid receptor regulation of striatal dopamine D2-receptor mediated transmission. Sci Rep 6:37834.

Mamaligas AA, Barcomb K, Ford CP (2019) Cholinergic transmission at muscarinic synapses in the striatum is driven equally by cortical and thalamic inputs. Cell Rep 28:1003–1014.e3.

Mateo Y, Johnson KA, Covey DP, Atwood BK, Wang HL, Zhang S, Gildish I, Cachope R, Bellocchio L, Guzman M, Morales M, Cheer JF, Lovinger DM (2017) Endocannabinoid actions on cortical terminals orchestrate local modulation of dopamine release in the nucleus accumbens. Neuron 96:1112–1126.e5.

Matsuda W, Furuta T, Nakamura KC, Hioki H, Fujiyama F, Arai R, Kaneko T (2009) Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. J Neurosci 29:444–453.

Melchior JR, Ferris MJ, Stuber GD, Riddle DR, Jones SR (2015) Optogenetic versus electrical stimulation of dopamine terminals in the nucleus accumbens reveals local modulation of presynaptic release. J Neurochem 134:833–844.

Quinn DM (1987) Acetylcholinesterase: enzyme structure, reaction dynamics, and virtual transition states. Chem Rev 87:955–979.

Quiroz C, Orrú M, Rea W, Ciudad-Roberts A, Yepes G, Britt JP, Ferré S (2016) Local control of extracellular dopamine levels in the medial nucleus accumbens by a glutamatergic projection from the infralimbic cortex. J Neurosci 36:851–859.

Rossi J, Balthasar N, Olson D, Scott M, Berglund E, Lee CE, Choi MJ, Lauzon D, Lowell BB, Elmquist JK (2011) Melanocortin-4 receptors expressed by cholinergic neurons regulate energy balance and glucose homeostasis. Cell Metab 13:195–204.

Schultz W, Apicella P, Ljungberg T (1993) Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. J Neurosci 13:900–913.

Shimizu N, Duan SM, Hori T, Oomura Y (1990) Glutamate modulates dopamine release in the striatum as measured by brain microdialysis. Brain Res Bull 25:99–102.

Shin JH, Adrover MF, Alvarez VA (2017) Distinctive modulation of dopamine release in the nucleus accumbens shell mediated by dopamine and acetylcholine receptors. J Neurosci 37:11166–11180.

Shin JH, Adrover MF, Wess J, Alvarez VA (2015) Muscarinic regulation of dopamine and glutamate transmission in the nucleus accumbens. Proc Natl Acad Sci USA 112:8124–8129.

Sulzer D, Cragg SJ, Rice ME (2016) Striatal dopamine neurotransmission: regulation of release and uptake. Basal Ganglia 6:123–148.

Surmeier DJ, Ding J, Day M, Wang Z, Shen W (2007) D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. Trends Neurosci 30:228–235.

Taber MT, Fibiger HC (1993) Electrical stimulation of the medial prefrontal cortex increases dopamine release in the striatum. Neuropsychopharmacology 9:271–275.

Taber MT, Fibiger HC (1994) Cortical regulation of acetylcholine release in rat striatum. Brain Res 639:354–356.

Threlfell S, Clements MA, Khodai T, Pienaar IS, Exley R, Wess J, Cragg SJ (2010) Striatal muscarinic receptors promote activity dependence of dopamine transmission via distinct receptor subtypes on cholinergic interneurons in ventral versus dorsal striatum. J Neurosci 30:3398–3408.

Threlfell S, Lalic T, Platt NJ, Jennings KA, Deisseroth K, Cragg SJ (2012) Striatal dopamine release is triggered by synchronized activity in cholinergic interneurons. Neuron 75:58–64.

Tritsch NX, Sabatini BL (2012) Dopaminergic modulation of synaptic transmission in cortex and striatum. Neuron 76:33–50.

Tritschler L, Kheirbek MA, Dantec YL, Mendez-David I, Guilloux JP, Faye C, Doan J, Pham TH, Hen R, David DJ, Gardier AM (2018) Optogenetic activation of granule cells in the dorsal dentate gyrus enhances dopaminergic neurotransmission in the nucleus accumbens. Neurosci Res 134:56–60.

Vidal CJ, Chai MS, Plummer DT (1987) The effect of temperature on the activity of acetylcholinesterase preparations from rat brain. Neurochem Int 11:135–141.

Voorn P, Vanderschuren LJ, Groenewegen HJ, Robbins TW, Pennartz CM (2004) Putting a spin on the dorsal-ventral divide of the striatum. Trends Neurosci 27:468–474.

Wang L, Shang S, Kang X, Teng S, Zhu F, Liu B, Wu Q, Li M, Liu W, Xu H, Zhou L, Jiao R, Dou H, Zuo P, Zhang X, Zheng L, Wang S, Wang C, Zhou Z (2014) Modulation of dopamine release in the striatum by physiologically relevant levels of nicotine. Nat Commun 5:3925.

Zhang H, Sulzer D (2004) Frequency-dependent modulation of dopamine release by nicotine. Nat Neurosci 7:581–582.

Zhou FM, Liang Y, Dani JA (2001) Endogenous nicotinic cholinergic activity regulates dopamine release in the striatum. Nat Neurosci 4:1224–1229.

Zhu J, Hexum TD (1992) Characterization of cocaine-sensitive dopamine uptake in PC12 cells. Neurochem Int 21:521–526.

frontiers in
HUMAN NEUROSCIENCE

REVIEW ARTICLE
published: 31 December 2013
doi: 10.3389/fnhum.2013.00904

More than a face: a unified theoretical perspective on nonverbal social cue processing in social anxiety

Eva Gilboa-Schechtman* and Iris Shachar-Lavie

Department of Psychology, The Gonda Brain Science Center, Bar-Ilan University, Ramat Gan, Israel

Edited by:
Quincy Wong, Macquarie University, Australia

Reviewed by:
Oliver C. Schultheiss, Friedrich Alexander University, Germany
Robert Edelmann, Univeristy of Roehampton, UK

****Correspondence:***
Eva Gilboa-Schechtman, Department of Psychology, The Gonda Brain Science Center, Bar-Ilan University, Building 902, Ramat Gan 52900, Israel
e-mail: evagilboa@gmail.com

Processing of nonverbal social cues (NVSCs) is essential to interpersonal functioning and is particularly relevant to models of social anxiety. This article provides a review of the literature on NVSC processing from the perspective of social rank and affiliation biobehavioral systems (ABSs), based on functional analysis of human sociality. We examine the potential of this framework for integrating cognitive, interpersonal, and evolutionary accounts of social anxiety. We argue that NVSCs are uniquely suited to rapid and effective conveyance of emotional, motivational, and trait information and that various channels are differentially effective in transmitting such information. First, we review studies on perception of NVSCs through face, voice, and body. We begin with studies that utilized information processing or imaging paradigms to assess NVSC perception. This research demonstrated that social anxiety is associated with biased attention to, and interpretation of, emotional facial expressions (EFEs) and emotional prosody. Findings regarding body and posture remain scarce. Next, we review studies on NVSC expression, which pinpointed links between social anxiety and disturbances in eye gaze, facial expressivity, and vocal properties of spontaneous and planned speech. Again, links between social anxiety and posture were understudied. Although cognitive, interpersonal, and evolutionary theories have described different pathways to social anxiety, all three models focus on interrelations among cognition, subjective experience, and social behavior. NVSC processing and production comprise the juncture where these theories intersect. In light of the conceptualizations emerging from the review, we highlight several directions for future research including focus on NVSCs as indexing reactions to changes in belongingness and social rank, the moderating role of gender, and the therapeutic opportunities offered by embodied cognition to treat social anxiety.

Keywords: social anxiety, non-verbal behavior, social rank, dominance, affiliation, production, expression

INTRODUCTION

Individuals with social anxiety disorder (SAD) suffer from significant functional impairment in multiple aspects of daily life: Their academic trajectory is frequently interrupted, they are employed below potential, and their social functioning suffers (Stein and Kean, 2000; Eng et al., 2005; Aderka et al., 2012). Individuals who are high in social anxiety perceive themselves as submissive and behave in nonassertive ways (Aderka et al., 2009; Schneier et al., 2009; Weeks et al., 2011). Moreover, social anxiety has been associated with lower reported quality of intimacy in peer, friendship, and romantic relationships (Rodebaugh, 2009; Cuming and Rapee, 2010; Weisman et al., 2011). Detailed analyzes of interpersonal behavior suggested that individuals with high social anxiety appear less confident, less affiliative, and less synchronous in social interactions than individuals with low social anxiety (e.g., Fydrich et al., 1998; Kachin et al., 2001; Alden and Taylor, 2004).

In this article we examine difficulties in interpersonal communication in social anxiety from the perspective of functional analysis of human sociality (e.g., Bugental, 2000). From this perspective, each individual participates in several evolutionarily shaped social structures. The two most prominent structures are affiliative relationships (friendship, companionship, intimacy) and hierarchical relationships (authority, social rank, social power). The need to affiliate with or belong to a social group is considered one of the central social motives across species, and basic psychological systems are hypothesized to constantly monitor for inclusionary status (e.g., Baumeister and Leary, 1995). Similarly, a need to advance in the social hierarchy and to be sensitive to threats to one's status within a group appears to be inherited from our primate ancestors (Sapolsky, 2005). Social exclusion (i.e., ostracism or social rejection) and social submission (e.g., being defeated) threaten one's belonging to and standing in a social group. Such events also decrease one's chances of future social effectiveness and collaboration. In contrast, social acceptance and social ascendance increase one's chances of social flourishing. In the following, we review some evidence that social rank and affiliation are basic psychological systems and then we examine their links with nonverbal social cue (NVSC).

SOCIAL RANK BIOBEHAVIORAL SYSTEM (SRBS)

A distinctive feature of social species' cooperative living is the formation of dominance hierarchies (e.g., Rowell, 1974). Conspecific members of the species face competition for resources, leading to aggressive interactions (West-Eberhard, 1979). Social hierarchies afford dominant members privileged access to food and mates, thereby conferring survival advantages. Such hierarchies are social systems containing an "implicit or explicit rank order of individuals or groups with respect to a valued social dimension" (Magee and Galinsky, 2008, p. 354). Social hierarchies contribute to stable social organization, and this stability reduces the costs of social competition for both dominant and subordinate members (e.g., Sloman and Armstrong, 2002). Possibly due to its importance for survival, a specialized biobehavioral system that monitors for social rank appears to have developed in humans and other mammals. This biobehavioral system has been called the rank regulation system (Zuroff et al., 2010), hierarchical domain (Bugental, 2000), power (Shaver et al., 2011), or dominance behavior system (Johnson et al., 2012).

The Social rank biobehavioral system (SRBS) is postulated as constantly monitoring one's standing in relation to others and using that information to guide behavior (e.g., Silk, 2007). Neuroimaging evidence supported the role of limbic, prefrontal, and striatal pathways (Beasley et al., 2012) and possibly intraparietal sulci (Chiao et al., 2009) in human social rank processing. Moreover, individual differences in social rank were linked to neural activation of limbic and frontal pathways when viewing social information (e.g., Muscatell et al., 2012). The most frequently studied SRBS-related biochemical substrate has been testosterone (e.g., Schultheiss and Wirth, 2008). Testosterone was found to correlate with self-report, observational, and cognitive measures of dominance in men and women alike (e.g., Archer, 2006; Sellers et al., 2007); additionally, estradiol correlated with dominance in females (e.g., Stanton and Schultheiss, 2007).

SRBS appears to emerge early in the developmental sequence (Thomsen et al., 2011), to operate automatically (Moors and De Houwer, 2005; Tracy et al., 2013) and fluently (Zitek and Tiedens, 2012), and to be specifically attuned to certain nonverbal signals such as gaze, voice, gestures, and postures (Wolff and Puts, 2010; Terburg et al., 2012; Tracy et al., 2013). Thus, SRBS is postulated to organize and orchestrate individuals' responses to changes in social standing.

AFFILIATION BIOBEHAVIORAL SYSTEM (ABS)

Over the course of human evolutionary history, members of the same species depended on each other for survival. In humans and other mammals, individuals who could gather support from their social surroundings gained access to more resources and therefore increased their chances of survival and reproduction. Like the dominance system, affiliative system is another basic biobehavioral system which continuously monitors for inclusionary status (Baumeister and Leary, 1995). Indeed, failing to satisfy the need to belong was found to activate neural circuits that partially overlap those of physical pain (e.g., Eisenberger et al., 2003; Dewall et al., 2010). Specifically, social exclusion was associated with greater activity in dorsal anterior cingulate cortex (dACC) and anterior insula. Again, individual differences in sensitivity to social exclusion (such as rejection sensitivity and attachment anxiety) were associated with greater responses in these regions (Linnen et al., 2012). Oxytocin and vassopressin, neuropeptides germane to affiliative behavior, may be implicated in the regulation of interpersonal stress (Taylor et al., 2000) and as affecting prosocial behavior (Poulin et al., 2012). Relatedly, intranasal administration of oxytocin reduced distress following social rejection in women who endorsed emotional coping strategies (Cardoso et al., 2012). Moreover, oxytocin administration increased prosocial behavior in women with a history of positive parenting (Riem et al., 2013). ABS emerges early in the developmental sequence (Feldman, 2012), operates automatically (Lakin et al., 2008), and is attuned to nonverbal signals of touch, gaze, and vocalization (Guastella et al., 2008; Dunbar and Abra, 2010; Farley et al., 2013).

Some convergence exists at the conceptual level between the close social bonds system and the belongingness system. In general, the jury is still out whether these two systems are best conceptualized as distinct (e.g., Panksepp and Watt, 2011), or form a single coherent unit (e.g., Feldman, 2012). For purposes of this review we focus on their similarities and view them as one. Thus, we see the ABS as organizing and orchestrating individuals' responses to opportunities to, and ruptures in, social bonds.

SOCIAL RANK, AFFILIATION, AND NONVERBAL SOCIAL CUES (NVSCs)

Consistent evidence showed that social rank is more frequently expressed through nonverbal than verbal cues (e.g., Argyle et al., 1970; Mehrabian, 1970; Mignault and Chaudhuri, 2003). Moreover, when both verbal and nonverbal cues of social rank were present, nonverbal cues were more likely to influence observers' judgments than verbal ones (e.g., Argyle et al., 1970; Jacob et al., 2012). Similarly, affection is also frequently expressed via nonverbal cues such as touch, vocalization, and gaze (e.g., Feldman and Eidelman, 2007; App et al., 2012; Farley et al., 2013). Indeed, people communicate emotions through multiple nonverbal channels such as face, body, voice, and touch (e.g., Buck, 1984).

Communication efficacy has been shown to vary across nonverbal channels in accruing information about likability, dominance, and trustworthiness (e.g., Zuckerman and Driver, 1989; Hall et al., 2005; Todorov et al., 2009, 2013). Researchers posited that information encoded in postures and voices is transmitted more effectively to larger audiences across longer distances, such as to one's social group or across social groups, whereas information encoded in face and touch is more effectively transmitted to proximate others (Tracy and Robins, 2004; App et al., 2012). Correspondingly, social status emotions (such as pride or shame) were shown to be communicated more effectively through the body than through face or touch (App et al., 2012). Moreover, vocal information was found to effectively communicate dominance (Wolff and Puts, 2010), and research on animal and human behavior alike has documented strong links between expansive body postures and trait and state dominance (Weisfeld and Beresford, 1982; Ellyson and Steve, 1985; Hall et al., 2005). Indeed, Stanton et al. (2010) have postulated that the meaning of nonverbal signals as motivated and rewarding for the sender and the perceiver, respectively. Altogether, different lines of research converge in suggesting that: (a) NVSCs are uniquely suited to rapid, effective transmission of emotional, motivational, and trait

information and (b) various channels differ in their effectiveness for transmitting this information.

SOCIAL ANXIETY AND NONVERBAL SOCIAL CUES (NVSCs)

To date, the main nonverbal signal examined in the social anxiety context was emotional facial expression (EFE; see Staugaard, 2010 for reviews). There are sound theoretical reasons to focus on EFEs. First, facial affect is instrumental in social development, emotion regulation, and social functioning (e.g., Leppänen and Hietanen, 2001). Second, facial affect is processed by specialized networks within a particular circuit of brain structures, some of which function abnormally in social anxiety (Nakao et al., 2011). Third, as argued by Bistricky et al. (2011), direct gaze can initiate automatic self-referent processing and self-relative-to-other processing (e.g., "Is his response to me unfavorable?"), and these processes are known to be problematic in social anxiety. Fourth, facial affect represents particularly salient information in close interpersonal situations (App et al., 2012).

Other NVSCs share several important characteristics with EFEs. First, NVSCs are mostly involuntary. As a result, they are more likely to serve as "honest signals" of individuals' internal emotions or attitudes toward an interaction partner than more strategically controlled verbal content. Correspondingly, social information processing should be better attuned to NVSCs than to verbally communicated interpersonal information (e.g., Gotlib et al., 2004). Thus, NVSCs may offer more "privileged" access to individuals' formation of approach and avoidance tendencies than would verbal expressions. Second, inasmuch as emotions are preparatory states for action, perceiving NVSCs should prime an immediate social reaction (Frijda, 1986). Indeed, some suggested that NVSCs may constitute a more evocative medium of transmitting emotions than verbal information (Buck and Vanlear, 2002). Third, given the primacy of NVSCs in the developmental sequence, they likely form the foundation for affective sensitivity and regulation (Cozolino, 2002). The preverbal foundation of implicit affective memories has been assumed to form a lasting basis for self and other schemas (e.g., Bistricky et al., 2011).

Although examinations of EFE processing have extended and deepened understanding of social anxiety's basic processes and biases, much may be learned from expanding the framework in two ways. First, the inclusion of new signals and modalities—especially posture and voice—may elucidate how socially anxious individuals process interpersonal information and engage in impression formation and revision. In the following, we seek to summarize the rather disperse literature on NVSC processing in social anxiety, supplementing it with the burgeoning literature on voice and posture processing among nonclinical populations.

Second, we seek to expand the established framework by highlighting the important yet relatively understudied area of NVSC expression. Thus far, the vast majority of studies have examined biases in perception of NVSCs. However, the expression (or production) of those cues is central to interpersonal behavior and may provide an unbiased measure of the expresser's emotional states and interpersonal tendencies and abilities. Moreover, expression of these cues may, in itself, affect the expresser's cognitive and emotional states, thereby attenuating or intensifying those states (e.g., Carney et al., 2010). Consequently, we will review literature on facial, vocal, and postural expression in social anxiety and nonclinical populations.

ORGANIZATION OF THE REVIEW

This review has three broad goals. First, we aim to systematically review empirical studies examining NVSC perception in social anxiety. We focus mostly on research that employed information-processing approaches while incorporating pertinent cognitive neuroscience findings. Specifically, we review findings on perception of faces (including gaze orientation), voices, and bodies (postures and gestures). Second, we aim to review literature on NVSC production. Here we concentrate on mimicry of facial expression, eye gaze, vocal productions, and adoption of posture and body movements. Due to the paucity of research on some of these topics, we will draw on research with normal populations. Third, we aim to integrate these findings with the three prominent theories of social anxiety, and to formulate new testable hypotheses that may contribute to better understanding of this condition's basic underlying mechanisms.

PERCEPTION OF NONVERBAL SOCIAL CUES (NVSCs)

Theoretical accounts converge in suggesting that misinterpretations of neutral or affiliative social signals as threatening are likely to deepen distress and contribute to the maintenance of SAD (e.g., Clark and Wells, 1995; Rapee and Heimberg, 1997; Gilbert, 2001; Alden and Taylor, 2004; Hofmann et al., 2004). Consequently, enhanced understanding of factors that influence biased or inaccurate interpretations of NVSC signals may help formulate a more complete, accurate model of SAD. We review available evidence of biases in processing of cues from faces, voices, and bodies. For each channel, we first review the perception of emotional stimuli (i.e., emotional facial expression, prosody, emotional gestures), followed by a review of impression formation from stable cues (e.g., facial features, basic vocal and body characteristics) and integrated nonverbal representations.

FACES

Emotion facial expression (EFE)

We first review studies focusing on EFEs with direct gaze and then studies involving variations in gaze direction.

Attention. In a comprehensive review, Staugaard (2010) concluded that biased processing of threatening vs. neutral or smiling EFEs tended to emerge mostly under conditions of brief exposure and to disappear when exposure time increased. Staugaard's review, which included mostly studies of EFE processing in non-stressful conditions, yielded only elusive differences between individuals with high vs. low social anxiety in attentional processing of threatening EFEs. Overall, eye-tracking studies found that social anxiety tends to correlate with biased attention to threatening as well as smiling EFEs. In fact, several studies found that socially anxious individuals exhibited reduced total fixation time to all emotional stimuli (e.g., Chen et al., 2012). Moreover, although highly socially anxious individuals revealed difficulty disengaging from threatening faces (e.g., Buckner et al., 2010; Moriya and Tanno, 2011; Schofield et al., 2012), they also exhibited slower attentional disengagement

from smiling, as compared to neutral, faces (Gilboa-Schechtman et al., 1999). However, using a continuous flickering paradigm, Wieser et al. (2011) found that angry compared to smiling and neutral expressions were associated with greater electrocortical facilitation over visual areas in individuals with high social anxiety, but not in those with low social anxiety.

Kolassa and Miltner (2006) found that angry expressions during an emotion identification task elicited enhanced right temporoparietal N170 in individuals with SAD (Kolassa and Miltner, 2006) and in subclinically socially anxious individuals (Mühlberger et al., 2009). The research team also reported generally enhanced P100 amplitudes both in individuals with SAD and with other anxiety disorders (Kolassa et al., 2007). However, amplitudes of the N170 component and later event related potentials (ERP) components did not differ between SAD, spider phobic, and control participants when schematic emotional faces were presented (Kolassa et al., 2007) or with morphed expressions (Kolassa et al., 2009). Recently, Peschard et al. (2013) identified an enhanced P1 component in processing emotional and neutral faces in social anxiety. Combined, these data suggest that social anxiety may be related to a specific temporal pattern in processing naturalistic facial expressions.

Using fMRI, several studies employing block-designs found enhanced amygdala reactivity to angry faces (Stein, 2002; Straube et al., 2005; Phan et al., 2006; Evans et al., 2008), neutral faces (Birbaumer et al., 1998; Veit et al., 2002), and happy faces (Straube et al., 2005), pointing to hyperactivation of this area in social anxiety for all emotional expressions. Furthermore, researchers suggested that amygdala activation in highly socially anxious individuals may depend more on EFE intensity than on the particular emotion expressed (Yoon et al., 2007).

Altogether, these studies suggest that individuals high in social anxiety exhibit enhanced processing of all facial expressions in some tasks, disengagement difficulties from both threatening and smiling expression (as compared to neutral expressions) on other tasks, and enhanced vigilance/reactivity to angry expression on select tasks.

Interpretation and evaluation. Many studies found no association between social anxiety and labeling accuracy in emotional labeling tasks, especially when participants had unlimited time for completion (e.g., Joormann and Gotlib, 2006; Arrais et al., 2010; Campbell et al., 2009; Heuer et al., 2010). Similarly, several rating studies did not identify differences between individuals with high vs. low social anxiety in evaluating single discrete EFEs (Stein et al., 2002) or mixed displays of smiling, neutral, and angry expressions (Gilboa-Schechtman et al., 2005; Lange et al., 2011). Other studies that examined response latencies to morphed or degraded presentations of emotional expressions typically showed that social anxiety correlated with a lower threshold for identifying angry expressions (e.g., Joormann and Gotlib, 2006; Gilboa-Schechtman et al., 2008). Recently, Arrais et al. (2010) reported that, compared to women low in social anxiety, women high in social anxiety (but not men) required less emotional information to identify smiling, sad, and fearful expressions. Interestingly, when time constraints were introduced in labeling studies, biased interpretation/evaluation of EFEs emerged (e.g., Heuer et al., 2010). In addition, Heuer et al. (2007) found that, as compared to low socially anxious individuals, high socially anxious individuals showed stronger avoidance tendencies of smiling and angry faces, while no such differences with respect to neutral faces were identified. Combined, these studies suggest that persons with high social anxiety tend to exhibit more negative evaluations of ambiguous or smiling EFEs, either on implicit evaluation tasks, or under conditions of high task difficulty. Additionally, when rating studies required participants to engage in impression formation or in predicting emotional impact of possible interactions, persons with high social anxiety rated smiling faces as less approachable and angry faces as more costly (e.g., Campbell et al., 2009; Douilliez et al., 2012).

Memory. The social anxiety literature is mixed regarding memory biases for experimentally presented stimuli. Some studies documented enhanced processing of threatening stimuli (e.g., Foa et al., 2000), whereas others found no such biases (e.g., Rapee et al., 1994; Coles and Heimberg, 2002; Rinck and Becker, 2005 for reviews). Other studies supported the erosion of positive memory biases in SAD (Liang et al., 2011). In general, however, the support is rather modest for negative memory bias or diminished positive bias in SAD using EFEs.

Gaze direction

Eye gaze plays an essential role in social interactions, as averted gaze may relay various social intentions such as submission (Mazur and Booth, 1998), disinterest (Itier and Batty, 2009), or even rejection (e.g., Wirth et al., 2010). Generally, direct gaze (vs. averted) was linked with observers' higher levels of physiological arousal (i.e., skin conductance), an effect enhanced for smiling faces (Pönkänen and Hietanen, 2012). In another study, participants rated averted gaze as less pleasant than direct gaze (Schmitz et al., 2012). In a virtual reality study, women high in social anxiety exhibited greater increase in avoidance while responding to male avatars when avatars had direct gaze than averted gaze while women with low social anxiety did not exhibit this pattern (Wieser et al., 2010). Similarly, high but not low social anxiety in women was linked with greater heart rate acceleration in response to direct, as compared to averted gaze (Wieser et al., 2009). Highly socially anxious individuals demonstrated avoidance of angry facial expressions but only when gaze was direct (vs. averted); however, they exhibited avoidance of smiling expressions regardless of gaze direction (Roelofs et al., 2010). Furthermore, an fMRI study confirmed preferential activation of brain areas related to fear response when SAD patients viewed direct gaze vs. averted gaze (Schneier et al., 2009).

It appears that sensitivity to EFEs may be modulated by gaze direction. Overall, direct (rather than averted) gaze results in higher physiological and neurological responsiveness as well as more pronounced avoidance among individuals high on social anxiety than those low in social anxiety. This sensitivity appears to be accentuated by emotional expressions; yet, more studies are needed to understand the nature of expression-gaze interaction. For example, depending on the emotional expression, direct gaze may suggest aggressive/dominant intent (when paired with angry

or even neutral facial expressions) or affiliative intent (when paired with smiling).

To summarize, the EFE literature suggests that in no-stress conditions social anxiety is associated with generalized reactivity to emotional and neutral faces alike. In addition, in some but not all tasks, social anxiety is associated with selective processing of threatening EFEs, and this bias appears to be modulated by the direction of the targets' gaze. Moreover, smiling facial expressions sometimes elicit reactions similar to those elicited by threatening EFEs. The latter finding highlights the complexity of examining facial expressions because they might simultaneously connote dominance and affiliation (Knutson, 1996). Thus, smiles could be negatively interpreted as threatening by virtue of their association with dominance. Alternatively, affiliative signals may invoke expectations of reciprocity, triggering self-evaluative concerns, which in turn can lead to fear and avoidance. The literature may benefit from better understanding how EFE processing is affected by contextual variations (such as gaze, head tilt, or gender) as well as by the perceiver's motivational states.

VOICE

Vocal information is an important channel for assessing an interaction partner's emotional states, intentions, and trait predispositions (Scherer, 1981; Siegman, 1987). Voices are processed early (e.g., Sauter and Eimer, 2010) and automatically (Donohue et al., 2012). Given the importance of vocal communication to the interpretation of social messages, it is surprising that only a handful of studies examined biased processing of vocal information in social anxiety. Existing studies involving social anxiety focus on labeling of, and brain responses to, emotional prosody of various utterances.

In the first study of its kind, Quadflieg et al. (2007) asked individuals with SAD and control participants to label the emotions of pseudo-words pronounced in angry, sad, happy, fearful, disgusted, or neutral tones by male and female actors. Compared to controls, individuals with SAD were more likely to label utterances as fearful or sad. Importantly, like in the evaluation of facial expressions, no group differences emerged regarding valence and arousal ratings for any of the emotional utterances.

In a follow-up study, individuals with SAD and matched controls labeled the emotion or gender of words pronounced with angry or neutral prosody by male and female actors during fMRI scanning (Quadflieg et al., 2008). Angry prosody elicited stronger brain activation than neutral prosody in limbic (insula, amygdala) as well as cortical areas for all participants. Importantly, compared to controls, individuals with SAD had increased activation in the right orbitofrontal cortex in response to angry vs. neutral voices under both task conditions. These results again substantiate findings for EFEs. McClure and Nowicki (2001) examined interpretation of EFE and vocal prosody in children who were high or low on social anxiety. Social anxiety was significantly associated with greater confusion in emotional labeling of vocal cues, where the high anxiety group was more likely than their less anxious peers to confuse other children's sad and fearful voices.

Taken together, the study of vocal prosody suggests that social anxiety appears to affect the interpretation of prosody characteristics. However, to the best of our knowledge, no studies to date examined attention to, or memory for, prosodic information. Moreover, biases in interpreting affectionate prosody remain to be explored.

BODY AND POSTURE

In the last decade, research on facial and vocal displays of emotions has been augmented by the study of postural expressions (e.g., Tracy and Robins, 2004; App et al., 2011; Dael et al., 2012). Indeed, posture appears particularly important for portraying and recognizing affective states. Posture and associated body language serve as a rich source of information for revealing others' goals, intentions, and emotions. These signals are visible and interpretable from a distance, do not require close contact, and allow simultaneous transmission of information to multiple individuals (e.g., Reed et al., 2003, 2006). The ability to engage in configural body processing seems to develop as early as 3 months of age (Gliga and Dehaene-Lambertz, 2005). Fearful, angry, and happy postures can be processed without visual awareness (Stienen and de Gelder, 2011). In addition, postures also transmit signals pertaining to individuals' social rank or status (e.g., Shariff and Tracy, 2009). For example, pride is cross-culturally recognized and easily distinguished from other similar emotions (e.g., happiness; Tracy and Robins, 2007). Recently, Rule et al. (2012) found that perceivers' judgments of posture dominance were significantly more accurate than chance guessing for exposures as brief as 40 ms, with no significant increase in accuracy given additional viewing time.

Few studies have examined processing of postural or any bodily information in social anxiety. Pitterman and Nowicki (2004) found that adults' correct labeling of standing postures (depicting happiness, fear, anger, and sadness) correlated negatively with fear of negative evaluation. In addition, compared to children with attention disorders, children high in social anxiety made more errors identifying angry posture in adult posers (Walker et al., 2011). De Gelder and her colleagues examined how negative affectivity and social inhibition associate with the processing of threatening body expressions (fear and anger; Kret et al., 2011). They found that negative affectivity correlated with deactivation of the core emotion system (e.g., amygdala, right insula), whereas social inhibition correlated with a tendency to activate a broad cortical network (e.g., orbitofrontal cortex). Although tentative, these results suggest that social anxiety may be associated with enhanced processing of threatening or dominant postures.

STABLE FEATURES OF FACES, VOICES, AND BODIES

Another line of research explores the perception of stable characteristics (such as facial symmetry, vocal characteristics, and body size) as indicative of various psychological features (Tinlin et al., 2013). Substantial evidence indicates that people instantly form impressions from facial characteristics (e.g., Bar et al., 2006) and that these impressions affect important decisions (e.g., Olivola and Todorov, 2010). Regarding faces, individuals' facial maturity was linked to their perception as dominant. When baby-faced targets were paired with submissive information (congruent condition), they were better remembered than when they were paired

with dominant (non-congruent) information (Cassidy et al., 2012). In addition, increase in men's face-width-to-height ratio was linked to their faces being perceived as more aggressive (Carré et al., 2010). Thus, stable facial characteristics appear influential in judgments of people as dominant and/or affiliative.

Studies with non-clinical populations have examined the influence of stable auditory characteristics of voice—especially pitch—on trait judgments. Specifically, male participants made judgments of likely physical aggressiveness based on vocal recordings of men reading a sentence, which were raised or lowered in both fundamental frequency (F0) and formant dispersion (df; Wolff and Puts, 2010). Raised vocal masculinity (lowered F0 or df) yielded higher dominance ratings. Moreover, raters with either high or low levels of testosterone rated other men as more dominant than raters with mean levels of testosterone.

Stable characteristics of body may also influence perceptions of power and dominance. Height is an important factor, as taller individuals were perceived as more dominant (Adams, 1980; Melamed, 1992; Young and French, 1998), even when females were evaluated (Boyson and Butler, 1999).

Recent studies from our group have shown that social anxiety is associated with biased impression formation and impression revision from trait descriptions (Aderka et al., 2013; Haker et al., 2013). Yet, to date, studies have yet to assess the role of social anxiety in the interpretation of static characteristics of the faces, voices, or bodies. Such an endeavor may enrich understanding of how social anxiety affects person perception.

INTEGRATED NONVERBAL REPRESENTATIONS

Integrating information from face, voice, and body plays a crucial role in human ability to understand social interactions. Although distinct, these different sources of information are normally processed in parallel, and, when synchronous, facilitate social comprehension. In contrast, the study of channel incongruity may enable understanding of the relative diagnostic value of each component of person perception. For example, participants predicted that the face would be most influential on perceptions of winning and losing emotions; however, the body was the best diagnostic feature (Aviezer et al., 2012). Rule et al. (2012) also found that face-body gestalt increased accuracy in perceiving dominance expressions from bodies and faces. Van den Stock et al. (2007) demonstrated that body postures conveying emotions influenced recognition of facial expressions and tones of voice. These findings emphasize the importance of emotional whole-body expressions, as well as the combination of NVSCs in everyday settings. With respect to social anxiety, the examination of conflicting cues—such as those connoting threat in one channel and affiliativeness in another—may elucidate the relative diagnosticity of each information type for socially anxious individuals.

EVALUATIVE SUMMARY OF NONVERBAL SOCIAL CUE (NVSC) PERCEPTION

Taken together, the research reviewed above suggests that individuals with high social anxiety are more likely than peers with low social anxiety to misinterpret EFEs and vocal expressions of emotions, although results pinpointing threatening misinterpretations were not as robust as theoretically expected. While pertinent studies on body posture are scarce, it appears likely that socially anxious individuals may be also biased in processing these stimuli. Examination of socially anxious individuals' perceptions of nonverbal representations of social status cues is important, given that social status is central to evolutionary (e.g., Gilbert and Trower, 2001), interpersonal (Alden and Taylor, 2004), and cognitive (e.g., Clark and Wells, 1995) accounts of social anxiety.

PRODUCTION OF NONVERBAL SOCIAL CUES (NVSCs)

Existing research has focused on NVSC perception, but multiple considerations indicate the need to extend investigation to the domain of NVSC production. First, examination of expressive indices of emotion can substantially complement subjective reports by linking research on adult humans to research on infants (Cappella, 1981) and non-human primates (e.g., Geerts and Brüne, 2009). Second, expressivity was pinpointed as an important clue for judging cooperation (e.g., Boone and Buck, 2003; Schug et al., 2010) and thus may direct an interaction's outcome. Third, based on embodied cognition accounts, researchers found that expressive behaviors lead to cognitive change (Briñol et al., 2011). These behavior-cognition and cognition-behavior links may generate either a self-enforcing positive or negative cycle, again influencing the course of interactions. Finally, NVSC production was shown to predict the long-term course of depression following treatment (Bos et al., 2006, 2007). Thus, NVSC production may possibly also be used to predict treatment outcomes in social anxiety.

FACES

Facial mimicry

Mimicry has been conceptualized as affiliative social behavior where one emulates another person's nonverbal actions (Lakin and Chartrand, 2003). By and large, mimicry was found to generate positive social feelings (Chartrand and Bargh, 1999). For example, priming individuals with prosocial (as opposed to antagonistic) goals generated more mimicking behavior (Leighton et al., 2010). Also, a priori manipulation of liking for another increased mimicry (Stel et al., 2010). Furthermore, automatic mimicry responses increased following a social exclusion situation, possibly because the threat of exclusion promoted affiliative motivations and actions (Lakin et al., 2008).

Importantly, mimicry depends on an interaction's nature; competitive interactions seem to elicit less facial mimicry than collaborative interactions (Likowski et al., 2011). Lower mimicry in competitive situations may be conceptualized as complementary behavior, where dominant displays (e.g., anger) elicit submissive reactions (e.g., fear). Interestingly, mimicry not only affects this social signal's recipient but also it's expresser: women were slower to recognize the affective valence of briefly displayed facial expressions when constrained from mimicking them (Stel and van Knippenberg, 2008). This effect was attributed to the fact that facial constraints hinder women's capacity to empathize.

Only a handful of studies explored automatic facial mimicry in social anxiety. In the first study to address this issue, Dimberg (1997) found that women high in public speaking fearfulness reacted with more frowning to angry faces than did women low in public speaking fearfulness, and women low in public speaking

fearfulness exhibited more reactivity to happy facial expression than did women high in public speaking fearfulness (Dimberg, 1997). Vrana and Gross (2004) found that individuals high in public speaking fear exhibited more frowning in response to angry, neutral and happy facial expressions than did individuals low in speech fearfulness. Finally, Dimberg and Thunberg (2007) found that the individuals in a high public speaking fear group exhibited greater negative facial reactivity when responding to angry vs. happy faces, and greater positive emotional reactivity when responding to happy, as compared to angry faces. Given the recent emphasis on importance of mimicry for the establishment of rapport and liking on the one hand, and the deficits in affiliative behavior in social anxiety on the other (e.g., Alden and Taylor, 2011), more studies examining mimicry in social anxiety are needed.

Facial expressivity: voluntary displays of affect

Emotional expressivity is the extent to which an individual manifests emotional impulses behaviorally (Gross and John, 1997). Emotional expressivity was linked to increased positive affect (Gross and John, 1995; Burgin et al., 2012), whereas suppressing emotional displays was shown to decrease positive affect (Gross, 1998). Emotional expressivity was also linked to better social functioning (Burgin et al., 2012), cooperativeness (Schug et al., 2010), and agreeableness (Gross and John, 1995). Moreover, expressivity is beneficial not only when it concerns positive emotions. Feinberg et al. (2012) found that people who expressed embarrassment were judged as more prosocial. In terms of dominance, Hall et al. (2005) found that more gazing, nodding, and smiling and a generally more expressive face were associated with high interpersonal power.

Social anxiety is linked to lower levels of emotional expressivity (Kashdan and Breen, 2008), and is found to be mediated by beliefs that overt emotional expression is negative (Spokas et al., 2009). Moreover, Kashdan et al. (2011) found that the negative correlation between social anxiety and positive affect was mediated by the tendency to suppress emotional displays.

In sum, expressivity emerges as an important nonverbal variable in social interactions, signaling both dominant and affiliative propensities. Social anxiety is associated with lower levels of expressivity, constituting a potential deficiency in both social domains.

Eye gaze

Robust evidence suggests that avoidance of eye contact may be a central characteristic of SAD (e.g., Schneier et al., 2011). Specifically, fear and avoidance of eye contact consistently correlate with overall severity of social anxiety (e.g., Safren et al., 1999; Baker and Edelmann, 2002; Stein et al., 2004), and were recently found to decrease with successful treatment (e.g., Schneier et al., 2011). Studies relying on independent observers' judgments revealed that individuals with SAD made less eye contact during social interactions than individuals low in social anxiety (Baker and Edelmann, 2002; Voncken and Bögels, 2008). Advancements in eye-tracking methodology have permitted in-depth examination of this feature of socially anxious behavior: Individuals with SAD showed fewer gaze fixations on the eyes when facial expressions were presented for relatively long time intervals (e.g., Horley et al., 2004; Moukheiber et al., 2010, 2012; but see also Wieser et al., 2009).

Theoretically, gaze avoidance has been linked to submissive behavior in a variety of species (e.g., Mazur and Booth, 1998). Inasmuch as human social anxiety has been postulated as related to such submissive behaviors (e.g., Gilbert, 2001), perhaps gaze processing in social anxiety holds promise for creating a neurobiological marker of this disorder.

VOICE

Individuals with SAD are often concerned about showing auditory signs of anxiety during social performance (e.g., Hirsch and Clark, 2007). Early studies seeking to examine the veracity of these concerns tested the vocal performance of socially anxious individuals in interpersonal situations. Such highly socially anxious individuals were often rated by others as less competent in their vocal communication compared to controls (e.g., Borkovec et al., 1974; Fydrich et al., 1998; see review in Baker and Edelmann, 2002). Further studies highlighted relations between social anxiety and temporal features of spontaneous speech. For example, highly socially anxious and clinically distressed individuals with SAD paused more often and for longer durations, demonstrated slower speech rate, and evidenced restricted verbal output (e.g., Borkovec et al., 1973; Lewin et al., 1996; Hofmann et al., 1997).

In a pioneering study, Laukka et al. (2008) used acoustic analysis to explore the effect of social anxiety on objectively defined auditory parameters. Public speech samples of individuals with SAD were recorded preceding and following pharmacological intervention. Participants who reported lower anxiety following treatment demonstrated post-treatment decreases in mean F0 (subjectively perceived as pitch) and decreased proportions of silent pauses. In another study, Weeks et al. (2011) placed socially anxious men in a competitive interaction with another man over the positive attention of a female peer. Consistent with evolutionary predictions, highly socially anxious men manifested increased mean F0, whereas men low in social anxiety showed the opposite trend. Recently, Weeks et al. (2012) also compared acoustic characteristics of public speaking between a group with SAD and non-anxious controls. Males with SAD evidenced greater F0 in comparison to non-anxious individuals across both studies. For females, the inverse correlation between social anxiety and F0 was significant only when examined in patients with generalized SAD, and in response to in vivo social exposures. Importantly, gender-specific thresholds for mean F0 demonstrated excellent differentiation between patients with generalized SAD and non-anxious controls.

Galili et al. (2013) examined college students' vocal characteristics as a function of social anxiety by asking participants to record neutral, command, and request sentences and then analyzing these utterances' acoustic properties (mF0, intensity, speech rate, speech fluency). Social anxiety was associated with higher mF0 in men and women and with lesser vocal intensity in men. Moreover, compared to neutral sentences, social anxiety was associated with lesser increase of vocal intensity in command utterances, and with greater decrease of vocal intensity in request utterances. In men but not women, social anxiety also correlated

with slower speech rate in request sentences. Taken together, these results pinpoint F0 as a promising biobehavioral marker of social anxiety (at least in men). Moreover, vocal intensity, speech rate, and speech fluency also seem likely to be affected by social anxiety in socially stressful situations.

Other nonverbal parameters of speech such as volubility (time spent talking; Mast, 2002; Brescoll, 2012), successful interruptions (Farley, 2008), vocal expressivity (e.g., Dunbar and Abra, 2010), initiation of speech acts, and laughter (Gifford and Hine, 1994) have been linked with high interpersonal power. Likewise, volubility was the best predictor of observer-rated social performance in an interaction task (Stevens et al., 2010). Similar findings emerged in an impromptu speech task, where socially phobic patients exhibited significantly less volubility than individuals low in social anxiety (Beidel et al., 2010).

In sum, different aspects of acoustic performance appear to link significantly with social anxiety. Acoustic parameters of speech associate closely with manifestations of power and social rank in both humans (e.g., Dunbar and Abra, 2010) and other mammals (e.g., Koren et al., 2008; Laporte and Zuberbühler, 2010). Altogether, these findings point to the possible diagnosticity of this expressive behavior.

POSTURE AND BODY MOVEMENT

Body posture and movement are important indicators of powerful behavior. Specifically, power is associated with more bodily openness, more erect or tense posture, and more body or leg shifts (Hall et al., 2005). Recently, Weeks et al. (2011) found that high social anxiety levels were associated with slumped and closed posture when interacting with a male competitor, whereas low social anxiety was associated with expansive posture. Other studies on observer ratings found that socially phobic individuals exhibited more bodily discomfort during social or performance tasks, such as rigidness and fidgeting (Voncken and Bögels, 2008; Heiser et al., 2009). Finally, measurement of observers' head movements showed that, when viewing an avatar's whole body, observers with high social anxiety mimicked the avatar significantly less than individuals with low social anxiety (Vrijsen et al., 2010).

EVALUATIVE SUMMARY OF NONVERBAL SOCIAL CUE (NVSC) PRODUCTION

Vocal and bodily behaviors appear linked to social anxiety and to expression of social rank. Facial expression and eye gaze seem linked both to expression of dominance (e.g., anger, contempt) and expression of affiliation (e.g., smiles). Existing research indicates that highly socially anxious individuals, especially males, will more likely exhibit submissive behaviors than individuals with low social anxiety.

Production of non-affiliative or submissive NVSCs is likely to deepen people's sense of disconnection and/or ineptitude either by directly influencing cognitions (e.g., Briñol et al., 2011) or through standards of culturally-appropriate behavioral norms (e.g., one should look one's offender in the eye, rather than lowering one's head; people should look others in the eye and smile). Thus, gaining more complete, complex understanding of NVSC production may advance more nuanced conceptualizations of interpersonal and situational factors that influence self-evaluations in social anxiety.

INTEGRATION OF THEORETICAL APPROACHES TO SOCIAL ANXIETY

NONVERBAL SOCIAL CUES (NVSCs) AT THE EPICENTER OF SOCIAL ENCOUNTERS

Independent theoretical contingents have identified different factors contributing to the maintenance of social anxiety: cognitions (e.g., Clark and Wells, 1995; Rapee and Heimberg, 1997; Gilbert, 2001; Hofmann et al., 2004), interpersonal factors (e.g., Alden and Taylor, 2004), or evolutionary pressures (Gilbert, 2001). Recent attempts were made to integrate these theoretical literatures. Our lab proposed a cognitive-evolutionary model (e.g., Aderka et al., 2013; Galili et al., 2013; Haker et al., 2013; Gilboa-Schechtman et al., in press), and several other researchers (Levinson et al., 2011; Taylor and Alden, 2011; Moscovitch et al., 2012) argued for a cognitive-interpersonal hypothesis. To varying degrees, the cognitive, interpersonal, and evolutionary models implicate abnormal social behavior and cognition in the onset and maintenance of social anxiety. Indeed, the three theories tend to be more complementary than contradictory. Importantly, all three postulate that socially anxious individuals are biased when perceiving social signals such as facial expressions, and because those signals guide interpersonal interactions, that bias leads to important outcomes. Interpersonal theories, evolutionary theories, and cognitive embodiment accounts also highlight the importance of the production of interpersonal signals (e.g., smiles, powerful postures) for the coordination of social interactions. Thus, the processing and production of NVSC is the juncture at which these theories intersect.

The evolutionary/interpersonal dimension of social rank and affiliation refine and sharpen the cognitive categories of threat and safety. Consider, for example, a case of an expanded pride posture (Tracy and Robins, 2004). While such posture may not represent social threat, it does connote an attempt to ascend in social rank. According to our account, such cues are likely to be selectively processed by individuals high in social anxiety. Conversely, the cognitive-embodiment account help explain the link between intra- and inter-personal mechanisms. For example, an assertive voice tone engenders a sense of self-assurance, which, in turn, elicits a complimentary (e.g., obliging, submissive) response from the interaction partner (e.g., Tiedens and Fragale, 2003). Moreover, we claim that the maintaining factors in social anxiety rely not only on the way individuals' process NVSC (the self-as-receiver of information) but also on the way individuals' *express* socially relevant attitudes and behavior (the self-as-an-agent).

REACTIVITY TO SOCIAL STRESS IN SOCIAL ANXIETY: NONVERBAL SOCIAL CUE (NVSC) AND BEYOND

Most theoretical models of SAD consider heightened sensitivity to, enhanced responsivity to, and impaired regulation in the face of social threat to be at the epicenter of this condition (e.g., Clark and Wells, 1995; Rapee and Heimberg, 1997; Gilbert and Trower, 2001; Hofmann et al., 2004). Individuals with SAD reported heightened emotional reactions to both positive and negative social events (e.g., Gilboa-Schechtman et al., 2000). Experimental

interpersonal manipulation studies consistently found that individuals with high social anxiety or SAD reported more intense, persistent negative affect in anticipation of, and following, various social challenges such as public speaking (e.g., Rapee and Lim, 1992; Ly and Roelofs, 2009), social ostracism (Oaten et al., 2008) and social success (Wallace and Alden, 1997) compared to individuals with low social anxiety.

Research has yet to examine the impact of social anxiety on the sensitivity to and bias in interpreting NVSCs following social stress. As our review suggests, two types of distinct social events appear crucial for such future examination: events indicating changes in belongingness (e.g., social exclusion, gaining social favor) and events indicating changes in social rank (e.g., defeat, victory). Examination of sensitivity to diverse stressors is pivotal, as sensitivity to interpersonal cues signaling affiliation or social rank is essential for smooth navigation in the interpersonal world, especially when one seeks to recover from a social misfortune or to capitalize on social success. For example, following exclusion, a failure to correctly identify potential friends or allies (vs. foes) is likely to hamper social reintegration, thereby increasing the threat of additional rejection. Similarly, the failure to assert oneself once a social group observes one's positive qualities or deeds may hamper one's chances of advancing in the social hierarchy. Finally, the mediating effects of social anxiety on the tendency to interpret NVSCs as affiliative or dominant may elucidate the causal structure of attunement difficulties in this population.

FUTURE DIRECTIONS

The present perspective calls for several directions of future research. First, the current unified perspective accentuates the need to study gender differences in perceiving and perhaps even more so in expressing NVSCs. Men and women face somewhat different threats in navigating groups; women face greater risk for exclusion, whereas men face greater risk for physical defeat (e.g., Archer, 2004; Benenson et al., 2011). Thus, we can expect gender to moderate the relations between social anxiety and perception of exclusion and dominance signals, where women are more sensitive to exclusion and men to dominance. Moreover, based on the tend-and-befriend theory of women's response to social threat, females may be expected to engage in more affiliative gestures following exclusion or defeat, whereas men may be more likely to express signals of deference or submissiveness (Taylor, 2006).

Second, existing evidence points to the possibility that embodiment of powerful NVSCs may lead to congruent changes in cognitions and cognitive processing and, conversely, the adoption of submissive NVSCs may deepen existing cognitive biases. For example, recent studies by Galinsky et al. (2006) suggested that merely thinking about powerful experiences enhances performance (Lammers et al., 2013). Moreover, both power postures and power roles were shown to reduce interpersonal fearfulness and increase approach behaviors; yet, power postures were more effective (Huang and Galinsky, 2010). Similarly, recent studies found that pitch does not merely correlate with social rank—it also instills subjective feelings of dominance and cognitive correlates of power (Stel et al., 2012). Altogether, these studies suggest reciprocal patterns of influence between NVSCs and cognition, with dominant expressions enhancing sense of power and submissive expressions decreasing it. Thus, loss of felt power may result in fewer displays of expressive behaviors in general and greater displays of submissive or appeasement behaviors (Stel et al., 2012). Such changes in expressive production may foster self-perceptions of powerlessness. These findings underscore an intriguing possibility for embodiment-focused cognitive interventions, such that socially anxious individuals may engage in training to implicitly or explicitly affect their sense of power or dominance.

Finally, research on NVSCs may provide converging evidence regarding the brain circuits engaged in social anxiety. For example, if processing of several types of NVSCs is found to involve overlapping circuitry (amygdala, insula, DLPFC), such brain areas may be associated with a special significance for social anxiety. More broadly, focusing on the social rank and affiliation systems may offer a new and helpful conceptual framework to examine brain mechanisms previously labeled approach-related and avoidance-related (e.g., Quirin et al., 2013; Terburg and van Honk, 2013).

CONCLUSION

NVSC are signals with long evolutionary history. As such they figured prominently in cognitive, interpersonal, and evolutionary accounts. In the present review we emphasized the links between biased processing of NVSC and social anxiety, highlighting the tendency of the socially anxious to be sensitive to social status cues. We also emphasized that production of non-affiliative or submissive NVSCs is likely to deepen a sense of disconnection and ineptitude. Given these propensities, socially anxious individuals might perceive and remember the social world as a hierarchical and competitive arena. These perceptions, in turn, may lead to a persistent sense of incompetence and inferiority (Gilbert et al., 2009). In the review we emphasized the interplay between perception and expression of NVSC as concomitant with social anxiety; future work would profit from examining the causal status of NVSC as contributing to the onset and maintenance of this condition.

REFERENCES

Adams, G. R. (1980). Social psychology of beauty: effects of age, height and weight on self-reported personality traits and social behavior. *J. Soc. Psychol.* 112, 287–293. doi: 10.1080/00224545.1980.9924330

Aderka, I. M., Haker, A., Marom, S., Hermesh, H., and Gilboa-Schechtman, E. (2013). Information-seeking bias in social anxiety disorder. *J. Abnorm. Psychol.* 122, 7–12. doi: 10.1037/a0029555

Aderka, I. M., Hofmann, S. G., Nickerson, A., Hermesh, H., Gilboa-Schechtman, E., and Marom, S. (2012). Functional impairment in social anxiety disorder. *J. Anxiety Disord.* 26, 393–400. doi: 10.1016/j.janxdis.2012.01.003

Aderka, I. M., Weisman, O., Shahar, G., and Gilboa-Schechtman, E. (2009). The roles of the social rank and attachment systems in social anxiety. *Pers. Individ. Dif.* 47, 284–288. doi: 10.1016/j.paid.2009.03.014

Alden, L. E., and Taylor, C. T. (2011). Relational treatment strategies increase social approach behaviors in patients with generalized social anxiety disorder. *J. Anxiety Disord.* 25, 309–318. doi: 10.1016/j.janxdis.2010.10.003

Alden, L. E., and Taylor, C. T. (2004). Interpersonal processes in social phobia. *Clin. Psychol. Rev.* 24, 857–882. doi: 10.1016/j.cpr.2004.07.006

App, B., McIntosh, D. N., Reed, C. L., and Hertenstein, M. J. (2011). Nonverbal channel use in communication of emotion: how may depend on why. *Emotion* 11, 603–617. doi: 10.1037/a0023164

App, B., Reed, C. L., and McIntosh, D. N. (2012). Relative contributions of face and body configurations: perceiving emotional state and motion intention. *Cogn. Emot.* 26, 690–698. doi: 10.1080/02699931.2011.588688

Archer, J. (2004). Sex differences in aggression in real-world settings: a meta-analytic review. *Rev. Gen. Psychol.* 8, 291–322. doi: 10.1037/1089-2680.8.4.291

Archer, J. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neurosci. Biobehav. Rev.* 30, 319–345. doi: 10.1016/j.neubiorev.2004.12.007

Argyle, M., Salter, V., Nicholson, H., Williams, M., and Burgess, P. (1970). The communication of inferior and superior attitudes by verbal and non-verbal signals. *Br. J. Soc. Clin. Psychol.* 9, 222–231. doi: 10.1111/j.2044-8260.1970.tb00668.x

Arrais, K. C., Machado-de-Sousa, J., Trzesniak, C., Santos Filho, A., Ferrari, M. C. F., Osório, F. L., et al. (2010). Social anxiety disorder women easily recognize fearfull, sad and happy faces: the influence of gender. *J. Psychiatr. Res.* 44, 535–540. doi: 10.1016/j.jpsychires.2009.11.003

Aviezer, H., Trope, Y., and Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science* 338, 1225–1229. doi: 10.1126/science.1224313

Baker, S. R., and Edelmann, R. J. (2002). Is social phobia related to lack of social skills? Duration of skill-related behaviours and ratings of behavioural adequacy. *Br. J. Clin. Psychol.* 41, 243–257. doi: 10.1348/014466502760379118

Bar, M., Neta, M., and Linz, H. (2006). Very first impressions. *Emotion* 6, 269–278. doi: 10.1037/1528-3542.6.2.269

Baumeister, R. F., and Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol. Bull.* 117, 497–529. doi: 10.1037//0033-2909.117.3.497

Beasley, M., Sabatinelli, D., and Obasi, E. (2012). Neuroimaging evidence for social rank theory. *Front. Hum. Neurosci.* 6:123. doi: 10.3389/fnhum.2012.00123

Beidel, D. C., Rao, P. A., Scharfstein, L., Wong, N., and Alfano, C. A. (2010). Social skills and social phobia: an investigation of DSM-IV subtypes. *Behav. Res. Ther.* 48, 992–1001. doi: 10.1016/j.brat.2010.06.005

Benenson, J. F., Markovits, H., Thompson, M. E., and Wrangham, R. W. (2011). Under threat of social exclusion, females exclude more than males. *Psychol. Sci.* 22, 538–544. doi: 10.1177/0956797611402511

Birbaumer, N., Grodd, W., Diedrich, O., Klose, U., Erb, M., Lotze, M., et al. (1998). fMRI reveals amygdala activation to human faces in social phobics. *Neuroreport* 9, 1223–1226. doi: 10.1097/00001756-199804200-00048

Bistricky, S. L., Ingram, R. E., and Atchley, R. A. (2011). Facial affect processing and depression susceptibility: cognitive biases and cognitive neuroscience. *Psychol. Bull.* 137, 998–1028. doi: 10.1037/a0025348

Boone, R., and Buck, R. (2003). Emotional expressivity and trustworthiness: the role of nonverbal behavior in the evolution of cooperation. *J. Nonverbal Behav.* 27, 163–182. doi: 10.1023/A:1025341931128

Borkovec, T. D., Fleischmann, D. J., and Caputo, J. A. (1973). The measurement of anxiety in an analogue social situation. *J. Consult. Clin. Psychol.* 41, 157–161. doi: 10.1037/h0035579

Borkovec, T. D., Stone, N. M., O'Brien, G. T., and Kaloupek, D. G. (1974). Evaluation of a clinically relevant target behavior for analog outcome research. *Behav. Ther.* 5, 503–513. doi: 10.1016/S0005-7894(74)80040-7

Bos, E. H., Bouhuys, A. L., Geerts, E., Van Os, T. W. D. P., and Ormel, J. (2006). Lack of association between conversation partners' nonverbal behavior predicts recurrence of depression, independently of personality. *Psychiatry Res.* 142, 79–88. doi: 10.1016/j.psychres.2005.05.015

Bos, E. H., Bouhuys, A. L., Geerts, E., Van Os, T. W. D. P., and Ormel, J. (2007). Stressful life events as a link between problems in nonverbal communication and recurrence of depression. *J. Affect. Disord.* 97, 161–169. doi: 10.1016/j.jad.2006.06.011

Boyson, A. R., and Butler, J. (1999). Height as power in women. *N. Am. J. Psychol.* 1, 109–114.

Brescoll, V. L. (2012). Who takes the floor and why: gender, power and volubility in organizations. *Adm. Sci. Quart.* 56, 622–641. doi: 10.1177/0001839212439994

Briñol, P., Petty, R., and Wagner, B. (2011). Embodied attitude change: a self validation perspective. *Soc. Personal.* 12, 1039–1050. doi: 10.1111/j.1751-9004.2011.00402.x

Buck, R. (1984). *The Communication of Emotion*. New York: Guilford Press.

Buck, R., and Vanlear, C. A. (2002). Verbal and nonverbal communication: distinguishing symbolic, spontaneous and pseudo-spontaneous nonverbal behavior. *J. Commun.* 52, 522–541. doi: 10.1111/j.1460-2466.2002.tb02560.x

Buckner, J. D., Maner, J. K., and Schmidt, N. B. (2010). Difficulty disengaging attention from social threat in social anxiety. *Cognit. Ther. Res.* 34, 99–105. doi: 10.1007/s10608-008-9205-y

Bugental, D. (2000). Acquisition of the algorithms of social life: a domain-based approach. *Psychol. Bull.* 126, 187–219. doi: 10.1037//0033-2909.126.2.187

Burgin, C. J., Brown, L. H., Royal, A., Silvia, P. J., Barrantes-Vidal, N., and Kwapil, T. R. (2012). Being with others and feeling happy: emotional expressivity in everyday life. *Pers. Individ. Dif.* 53, 185–190. doi: 10.1016/j.paid.2012.03.006

Campbell, D. W., Sareen, J., Stein, M. B., Kravetsky, L. B., Paulus, M. P., Hassard, S. T., et al. (2009). Happy but not so approachable: the social judgments of individuals with generalized social phobia. *Depress. Anxiety* 26, 419–424. doi: 10.1002/da.20474

Cappella, J. N. (1981). Mutual influence in expressive behavior: adult–adult and infant–adult dyadic interaction. *Psychol. Bull.* 89, 101–132. doi: 10.1037/0033-2909.89.1.101

Cardoso, C., Linnen, A.-M., Joober, R., and Ellenbogen, M. A. (2012). Coping style moderates the effect of intranasal oxytocin on the mood response to interpersonal stress. *Exp. Clin. Psychopharmacol.* 20, 84–91. doi: 10.1037/a0025763

Carney, D. R., Cuddy, A. J. C., and Yap, A. J. (2010). Power posing: brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychol. Sci.* 21, 1363–1368. doi: 10.1177/0956797610383437

Carré, J. M., Morrissey, M. D., Mondloch, C. J., and McCormick, C. M. (2010). Estimating aggression from emotionally neutral faces: which facial cues are diagnostic? *Perception* 39, 356–377. doi: 10.1068/p6543

Cassidy, B. S., Zebrowitz, L. A., and Gutchess, A. H. (2012). Appearance-based inferences bias source memory. *Mem. Cognit.* 40, 1214–1224. doi: 10.3758/s13421-012-0233-1

Chartrand, T. L., and Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910. doi: 10.1037//0022-3514.76.6.893

Chen, N. T. M., Clarke, P. J. F., MacLeod, C., and Guastella, A. J. (2012). Biased attentional processing of positive stimuli in social anxiety disorder: an eye movement study. *Cogn. Behav. Ther.* 41, 96–107. doi: 10.1080/16506073.2012.666562

Chiao, J. Y., Harada, T., Oby, E. R., Li, Z., Parrish, T., and Bridge, D. J. (2009). Neural representations of social status hierarchy in human inferior parietal cortex. *Neuropsychologia* 47, 354–363. doi: 10.1016/j.neuropsychologia.2008.09.023

Clark, D. M., and Wells, A. (1995). "A cognitive model of social phobia," in *Social Phobia: Diagnosis, Assessment and Treatment*, eds R. G. Heimberg, M. R. Liebowitz, D. A. Hope and F. R. Schneier (New York, NY: Guilford Press), 69–93.

Coles, M. E., and Heimberg, R. G. (2002). Memory biases in the anxiety disorders: current status. *Clin. Psychol. Rev.* 22, 587–627. doi: 10.1142/9789812776563_0041

Cozolino, L. (2002). *The Neuroscience of Psychotherapy: Building and Rebuilding the Human Brain*. New York, NY: WW Norton & Company.

Cuming, S., and Rapee, R. M. (2010). Social anxiety and self-protective communication style in close relationships. *Behav. Res. Ther.* 48, 87–96. doi: 10.1016/j.brat.2009.09.010

Dael, N., Mortillaro, M., and Scherer, K. R. (2012). Emotion expression in bosy action and posture. *Emotion* 12, 1085–1101. doi: 10.1037/a0025737

Dewall, C. N., Macdonald, G., Webster, G. D., Masten, C. L., Baumeister, R. F., Powell, C., et al. (2010). Acetaminophen reduces social pain: behavioral and neural evidence. *Psychol. Sci.* 21, 931–937. doi: 10.1177/0956797610374741

Dimberg, U. (1997). Social fear and expressive reactions to social stimuli. *Scand. J. Psychol.* 38, 171–174. doi: 10.1111/1467-9450.00024

Dimberg, U., and Thunberg, M. (2007). Speech anxiety and rapid emotional reactions to angry and happy facial expressions. *Scand. J. Psychol.* 48, 321–328. doi: 10.1111/j.1467-9450.2007.00586.x

Donohue, S. E., Liotti, M., Perez, R., and Woldorff, M. G. (2012). Is conflict monitoring supramodal? Spatiotemporal dynamics of cognitive control processes in an auditory Stroop task. *Cogn. Affect. Behav. Neurosci.* 12, 1–15. doi: 10.3758/s13415-011-0060-z

Douilliez, C., Yzerbyt, V., Gilboa-Schechtman, E., and Philippot, P. (2012). Social anxiety biases the evaluation of facial displays: evidence from single face and multi-facial stimuli. *Cogn. Emot.* 26, 1107–1115. doi: 10.1080/02699931.2011.632494

Dunbar, N. E., and Abra, G. (2010). Observations of dyadic power in interpersonal interaction. *Commun. Monogr.* 77, 657–684. doi: 10.1080/03637751.2010.520018

Eisenberger, N. I., Lieberman, M. D., and Williams, K. D. (2003). Does rejection hurt? An FMRI study of social exclusion. *Science* 302, 290–292. doi: 10.1126/science.1089134

Ellyson, L., and Steve, J. F. D. (1985). "Power, dominance and nonverbal behavior: basic concepts and issues," in *Power, Dominance and Nonverbal Behavior*, eds S. L. Ellyson and J. F. Davifio (New York: Springer), 1–27.

Eng, W., Coles, M. E., Heimberg, R. G., and Safren, S. A. (2005). Domains of life satisfaction in social anxiety disorder: relation to symptoms and response to cognitive-behavioral therapy. *J. Anxiety Disord.* 19, 143–156. doi: 10.1016/j.janxdis.2004.01.007

Evans, K. C., Wright, C. I., Wedig, M. M., Gold, A. L., Pollack, M. H., and Rauch, S. L. (2008). A functional MRI study of amygdala responses to angry schematic faces in social anxiety disorder. *Depress. Anxiety* 25, 496–505. doi: 10.1002/da.20347

Farley, S. D. (2008). Attaining status at the expense of likeability: pilfering power through conversational interruption. *J. Nonverbal Behav.* 32, 241–260. doi: 10.1007/s10919-008-0054-x

Farley, S. D., Hughes, S. M., and LaFayette, J. N. (2013). People will know we are in love: evidence of differences between vocal samples directed toward lovers and friends. *J. Nonverbal Behav.* 37, 123–138. doi: 10.1007/s10919-013-0151-3

Feinberg, M., Willer, R., and Keltner, D. (2012). Flustered and faithful: embarrassment as a signal of prosociality. *J. Pers. Soc. Psychol.* 102, 81–97. doi: 10.1037/a0025403

Feldman, R. (2012). Oxytocin and social affiliation in humans. *Horm. Behav.* 61, 380–391. doi: 10.1016/j.yhbeh.2012.01.008

Feldman, R., and Eidelman, A. I. (2007). Maternal postpartum behavior and the emergence of infant – mother and infant – father synchrony in preterm and full-term infants: the role of neonatal vagal tone. *Dev. Psychobiol.* 49, 290–302. doi: 10.1002/dev.20220

Foa, E. B., Gilboa-Schechtman, E., Amir, N., and Freshman, M. (2000). Memory bias in generalized social phobia: remembering negative emotional expressions. *J. Anxiety Disord.* 14, 501–519. doi: 10.1016/S0887-6185(00)00036-0

Frijda, N. H. (1986). *The Emotions*. New York, NY: Cambridge University Press.

Fydrich, T., Chambless, D. L., Perry, K. J., Buergener, F., and Beazley, M. B. (1998). Behavioral assessment of social performance: a rating system for social phobia. *Behav. Res. Ther.* 36, 995–1010. doi: 10.1016/s0005-7967(98)00069-2

Galili, L., Amir, O., and Gilboa-Schechtman, E. (2013). Acoustic properties of dominance and help seeking utterances in social anxiety. *J. Soc. Clin. Psychol.* 32, 651–673. doi: 10.1521/jscp.2013.32.6.651

Galinsky, A. D., Magee, J. C., Inesi, M. E., and Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychol. Sci.* 17, 1068–1074. doi: 10.1111/j.1467-9280.2006.01824.x

Geerts, E., and Brüne, M. (2009). Ethological approaches to psychiatric disorders: focus on depression and schizophrenia. *Aust. N.Z.J. Psychiatry* 43, 1007–1015. doi: 10.3109/00048670903270498

Gifford, R., and Hine, D. (1994). The role of verbal behavior in the encoding and decoding of interpersonal dispoitions. *J. Res. Pers.* 28, 115–132. doi: 10.1006/jrpe.1994.1010

Gilbert, P. (2001). Evolution and social anxiety: the role of attraction, social competition and social hierarchies. *Psychiatr. Clin. North Am.* 24, 723–751. doi: 10.1016/S0193-953X(05)70260-4

Gilbert, P., McEwan, K., and Bellew, R. (2009). The dark side of competition: how competitive behaviour and striving to avoid inferiority are linked to depression, anxiety, stress and self - harm. *Psychol. Psychother.* 82, 123–136. doi: 10.1348/147608308X379806

Gilbert, P., and Trower, P. (2001). "Evolution and process in social anxiety," in *International Handbook of Social Anxiety: Concepts, Research and Interventions Relating to the Self and Shyness*, eds W. R. Crozier and L. Alden (New York, NY, US: John Wiley & Sons Ltd), 259–279.

Gilboa-Schechtman, E., Foa, E., and Amir, N. (1999). Attentional biases for facial expressions in social phobia: the face-in-the-crowd paradigm. *Cogn. Emot.* 13, 305–318. doi: 10.1080/026999399379294

Gilboa-Schechtman, E., Foa, E., Vaknin, Y., Marom, S., and Hermesh, H. (2008). Interpersonal sensitivity and response bias in social phobia and depression: labeling emotional expressions. *Cognit. Ther. Res.* 32, 605–618. doi: 10.1007/s10608-008-9208-8

Gilboa-Schechtman, E., Franklin, M. E., and Foa, E. (2000). Anticipated reactions to social events: differences among individuals with generalized social phobia, obsessive compulsive disorder and nonanxious controls. *Cognit. Ther. Res.* 24, 731–746. doi: 10.1023/A:1005595513315

Gilboa-Schechtman, E., Presburger, G., Marom, S., and Hermesh, H. (2005). The effects of social anxiety and depression on the evaluation of facial crowds. *Behav. Res. Ther.* 43, 467–474. doi: 10.1016/j.brat.2004.03.001

Gilboa-Schechtman, E., Shachar, I., and Sahar, Y. (in press). "Positivity impairment as a broad-based feature of social anxiety," in *Handbook on Social Anxiety Disorder*, ed J. Weeks (New York: Wiley-Blackwell).

Gliga, T., and Dehaene-Lambertz, G. (2005). Structural encoding of body and face in human infants and adults. *J. Cogn. Neurosci.* 17, 1328–40. doi: 10.1162/0898929055002481

Gotlib, I. H., Krasnoperova, E., Yue, D. N., and Joormann, J. (2004). Attentional biases for negative interpersonal stimuli in clinical depression. *J. Abnorm. Psychol.* 113, 127–135. doi: 10.1037/0021-843x.113.1.121

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: divergent consequences for experience, expression and physiology. *J. Pers. Soc. Psychol.* 74, 224–237. doi: 10.1037//0022-3514.74.1.224

Gross, J. J., and John, O. P. (1995). Facets of emotional expressivity: three self-report factor and their correlates. *Pers. Individ. Dif.* 19, 555–568. doi: 10.1016/0191-8869(95)00055-b

Gross, J. J., and John, O. P. (1997). Revealing feelings: facets of emotional expressivity in self-reports, peer ratings and behavior. *J. Pers. Soc. Psychol.* 72, 435–448. doi: 10.1037/0022-3514.72.2.435

Guastella, A. J., Mitchell, P. B., and Dadds, M. R. (2008). Oxytocin increases gaze to the eye region of human faces. *Biol. Psychiatry* 63, 3–5. doi: 10.1016/j.biopsych.2007.06.026

Haker, A., Aderka, I. M., Marom, S., Hermesh, H., and Gilboa-Schechtman, E. (2013). Impression formation and revision in social anxiety disorder. *J. Anxiety Disord.* doi: 10.1016/j.janxdis.2013.05.001. [Epub ahead of print].

Hall, J., Coats, E. J., and LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. *Psychol. Bull.* 131, 898–924. doi: 10.1037/0033-2909.131.6.898

Heiser, N. A., Turner, S. M., Beidel, D. C., and Roberson-Nay, R. (2009). Differentiating social phobia from shyness. *J. Anxiety Disord.* 23, 469–476. doi: 10.1016/j.janxdis.2008.10.002

Heuer, K., Lange, W.-G., Isaac, L., Rinck, M., and Becker, E. S. (2010). Morphed emotional faces: emotion detection and misinterpretation in social anxiety. *J. Behav. Ther. Exp. Psychiatry* 41, 418–425. doi: 10.1016/j.jbtep.2010.04.005

Heuer, K., Rinck, M., and Becker, E. S. (2007). Avoidance of emotional facial expressions in social anxiety: the approach-avoidance task. *Behav. Res. Ther.* 45, 2990–3001. doi: 10.1016/j.brat.2007.08.010

Hirsch, C. R., and Clark, D. M. (2007). Imagery special issue: underestimation of auditory performance in social phobia and the use of audio feedback. *J. Behav. Ther. Exp. Psychiatry* 38, 447–458. doi: 10.1016/j.jbtep.2007.08.004

Hofmann, S. G., Gerlach, A. L., Wender, A., and Roth, W. T. (1997). Speech disturbances and gaze behavior during public speaking in subtypes of social phobia. *J. Anxiety Disord.* 11, 573–585. doi: 10.1016/s0887-6185(97)00040-6

Hofmann, S. G., Heinrichs, N., and Moscovitch, D. A. (2004). The nature and expression of social phobia: toward a new classification. *Clin. Psychol. Rev.* 24, 769–797. doi: 10.1016/j.cpr.2004.07.004

Horley, K., Williams, L. M., Gonsalvez, C., and Gordon, E. (2004). Face to face: visual scanpath evidence for abnormal processing of facial expressions in social phobia. *Psychiatry Res.* 127, 43–53. doi: 10.1016/j.psychres.2004.02.016

Huang, L., and Galinsky, A. D. (2010). No mirrors for the powerful: why dominant smiles are not processed using embodied simulation. *Behav. Brain Sci.* 33, 448. doi: 10.1017/s0140525x10001536

Itier, R. J., and Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neurosci. Biobehav. Rev.* 33, 843–863. doi: 10.1016/j.neubiorev.2009.02.004

Jacob, H., Kreifelts, B., Brück, C., Erb, M., Hösl, F., and Wildgruber, D. (2012). Cerebral integration of verbal and nonverbal emotional cues: impact of individual nonverbal dominance. *Neuroimage* 61, 738–747. doi: 10.1016/j.neuroimage.2012.03.085

Johnson, S. L., Leedom, L. J., and Muhtadie, L. (2012). The dominance behavioral system and psychopathology: evidence from self-report, observational and biological studies. *Psychol. Bull.* 138, 692–743. doi: 10.1037/a0027503

Joormann, J., and Gotlib, I. H. (2006). Is this happiness i see? Biases in the identification of emotional facial expressions in depression and social phobia. *J. Abnorm. Psychol.* 115, 705–714. doi: 10.1037/0021-843x.115.4.705

Kachin, K. E., Newman, M. G., and Pincus, A. L. (2001). An interpersonal problem approach to the division of social phobia subtypes. *Behav. Ther.* 32, 479–501. doi: 10.1016/s0005-7894(01)80032-0

Kashdan, T. B., and Breen, W. E. (2008). Social anxiety and positive emotions: a prospective examination of a self-regulatory model with tendencies to suppress or express emotions as a moderating variable. *Behav. Ther.* 39, 1–12. doi: 10.1016/j.beth.2007.02.003

Kashdan, T. B., Weeks, J. W., and Savostyanova, A. A. (2011). Whether, how and when social anxiety shapes positive experiences and events: a self-regulatory framework and treatment implications. *Clin. Psychol. Rev.* 31, 786–799. doi: 10.1016/j.cpr.2011.03.012

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *J. Nonverbal Behav.* 20, 165–182. doi: 10.1007/BF02281954

Kolassa, I.-T., Kolassa, S., Bergmann, S., Lauche, R., Dilger, S., Miltner, W. H. R., et al. (2009). Interpretive bias in social phobia: an ERP study with morphed emotional schematic faces. *Cogn. Emot.* 23, 69–95. doi: 10.1080/02699930801940461

Kolassa, I.-T., Kolassa, S., Musial, F., and Miltner, W. H. R. (2007). Event-related potentials to schematic faces in social phobia. *Cogn. Emot.* 21, 1721–1744. doi: 10.1080/02699930701229189

Kolassa, I.-T., and Miltner, W. H. R. (2006). Psychophysiological correlates of face processing in social phobia. *Brain Res.* 1118, 130–141. doi: 10.1016/j.brainres.2006.08.019

Koren, L., Mokady, O., and Geffen, E. (2008). Social status and cortisol levels in singing rock hyraxes. *Horm. Behav.* 54, 212–216. doi: 10.1016/j.yhbeh.2008.02.020

Kret, M. E., Denollet, J., Grèzes, J., and de Gelder, B. (2011). The role of negative affectivity and social inhibition in perceiving social threat: an fMRI study. *Neuropsychologia* 49, 1187–1193. doi: 10.1016/j.neuropsychologia.2011.02.007

Lakin, J. L., and Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychol. Sci.* 14, 334–339. doi: 10.1111/1467-9280.14481

Lakin, J., Chartrand, T. L., and Arkin, R. (2008). I am too just like you: nonconscious mimicry as an automatic behavioral response to social exclusion. *Psychol. Sci.* 19, 816–822. doi: 10.1111/j.1467-9280.2008.02162.x

Lammers, J., Dubois, D., Rucker, D. D., and Galinsky, A. D. (2013). Power gets the job: priming power improves interview outcomes. *J. Exp. Soc. Psychol.* 49, 776–779. doi: 10.1016/j.jesp.2013.02.008

Lange, W.-G., Heuer, K., Langner, O., Keijsers, G. P. J., Becker, E. S., and Rinck, M. (2011). Face value: eye movements and the evaluation of facial crowds in social anxiety. *J. Behav. Ther. Exp. Psychiatry* 42, 355–363. doi: 10.1016/j.jbtep.2011.02.007

Laporte, M. N. C., and Zuberbühler, K. (2010). Vocal greeting behaviour in wild chimpanzee females. *Anim. Behav.* 80, 467–473. doi: 10.1016/j.anbehav.2010.06.005

Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö., Faria, V., et al. (2008). In a nervous voice: acoustic analysis and perception of anxiety in social phobics' speech. *J. Nonverbal Behav.* 32, 195–214. doi: 10.1007/s10919-008-0055-9

Leighton, J., Bird, G., Orsini, C., and Heyes, C. (2010). Social attitudes modulate automatic imitation. *J. Exp. Soc. Psychol.* 46, 905–910. doi: 10.1016/j.jesp.2010.07.001

Leppänen, J. M., and Hietanen, J. K. (2001). Emotion recognition and social adjustment in school–aged girls and boys. *Scand. J. Psychol.* 42, 429–435. doi: 10.1111/1467-9450.00255

Levinson, C. A., Langer, J. K., and Rodebaugh, T. L. (2011). Self-construal and social anxiety: considering personality. *Pers. Individ. Dif.* 51, 355–359. doi: 10.1016/j.paid.2011.04.006

Lewin, M. R., McNeil, D. W., and Lipson, J. M. (1996). Enduring without avoiding: pauses and verbal dysfluencies in public speaking fear. *J. Psychopathol. Behav. Assess.* 18, 387–402. doi: 10.1007/bf02229142

Liang, C.-W., Hsu, W.-Y., Hung, F.-C., Wang, W.-T., and Lin, C.-H. (2011). Absence of a positive bias in social anxiety: the application of a directed forgetting paradigm. *J. Behav. Ther. Exp. Psychiatry* 42, 204–210. doi: 10.1016/j.jbtep.2010.12.002

Likowski, K. U., Mühlberger, A., Seibt, B., Pauli, P., and Weyers, P. (2011). Processes underlying congruent and incongruent facial reactions to emotional facial expressions. *Emotion* 11, 457–467. doi: 10.1037/a0023162

Linnen, A.-M., Ellenbogen, M. A., Cardoso, C., and Joober, R. (2012). Intranasal oxytocin and salivary cortisol concentrations during social rejection in university students. *Stress* 15, 393–402. doi: 10.3109/10253890.2011.631154

Ly, V., and Roelofs, K. (2009). Social anxiety and cognitive expectancy of aversive outcome in avoidance conditioning. *Behav. Res. Ther.* 47, 840–847. doi: 10.1016/j.brat.2009.06.015

Magee, J. C., and Galinsky, A. D. (2008). Social hierarchy: the self reinforcing nature of power and status. *Acad. Manag. Ann.* 2, 351–398. doi: 10.1080/19416520802211628

Mast, M. S. (2002). Dominance as expressed and inferred through speaking time. *Hum. Commun. Res.* 28, 420–450. doi: 10.1093/hcr/28.3.420

Mazur, A., and Booth, A. (1998). Testosterone and dominance in men. *Behav. Brain Sci.* 21, 353–363; discussion 363–397. doi: 10.1017/s0140525x98001228

McClure, E. B., and Nowicki, S. (2001). Associations between social anxiety and nonverbal processing skill in preadolescent boys and girls. *J. Nonverbal Behav.* 25, 3–19. doi: 10.1023/A:1006753006870

Mehrabian, A. (1970). A semantic space for nonverbal behavior. *J. Consult. Clin. Psychol.* 35, 248–257. doi: 10.1037/h0030083

Melamed, T. (1992). Personality correlates of physical height. *Pers. Individ. Dif.* 13, 1349–1350. doi: 10.1016/0191-8869(92)90179-s

Mignault, A., and Chaudhuri, A. (2003). The many faces of a neutral face: head tilt and perception of dominance and emotion. *J. Nonverbal Behav.* 27, 111–132. doi: 10.1023/A:1023914509763

Moors, A., and De Houwer, J. (2005). Automatic processing of dominance and submissiveness. *Exp. Psychol.* 52, 296–302. doi: 10.1027/1618-3169.52.4.296

Moriya, J., and Tanno, Y. (2011). The time course of attentional disengagement from angry faces in social anxiety. *J. Behav. Ther. Exp. Psychiatry* 42, 122–128. doi: 10.1016/j.jbtep.2010.08.001

Moscovitch, D. A., Rodebaugh, T. L., and Hesch, B. D. (2012). How awkward! Social anxiety and the perceived consequences of social blunders. *Behav. Res. Ther.* 50, 142–149. doi: 10.1016/j.brat.2011.11.002

Moukheiber, A., Rautureau, G., Perez-Diaz, F., Jouvent, R., and Pelissolo, A. (2012). Gaze behaviour in social blushers. *Psychiatry Res.* 200, 614–619. doi: 10.1016/j.psychres.2012.07.017

Moukheiber, A., Rautureau, G., Perez-Diaz, F., Soussignan, R., Dubal, S., Jouvent, R., et al. (2010). Gaze avoidance in social phobia: objective measure and correlates. *Behav. Res. Ther.* 48, 147–151. doi: 10.1016/j.brat.2009.09.012

Mühlberger, A., Wieser, M. J., Herrmann, M. J., Weyers, P., Tröger, C., and Pauli, P. (2009). Early cortical processing of natural and artificial emotional faces differs between lower and higher socially anxious persons. *J. Neural. Transm.* 116, 735–746. doi: 10.1007/s00702-008-0108-6

Muscatell, K. A., Morelli, S. A., Falk, E. B., Way, B. M., Pfeifer, J. H., Galinsky, A. D., et al. (2012). Social status modulates neural activity in the mentalizing network. *Neuroimage* 60, 1771–1777. doi: 10.1016/j.neuroimage.2012.01.080

Nakao, T., Sanematsu, H., Yoshiura, T., Togao, O., Murayama, K., Tomita, M., et al. (2011). fMRI of patients with social anxiety disorder during a social situation task. *Neurosci. Res.* 69, 67–72. doi: 10.1016/j.neures.2010.09.008

Oaten, M., Williams, K. D., Jones, A., and Zadro, L. (2008). The effects of ostracism on self–regulation in the socially anxious. *J. Soc. Clin. Psychol.* 27, 471–504. doi: 10.1521/jscp.2008.27.5.471

Olivola, C. Y., and Todorov, A. (2010). Elected in 100 milliseconds: appearance-based trait inferences and voting. *J. Nonverbal Behav.* 34, 83–110. doi: 10.1007/s10919-009-0082-1

Panksepp, J., and Watt, D. (2011). Why does depression hurt? Ancestral primary-process separation-distress (PANIC/GRIEF) and diminished brain reward (SEEKING) processes in the genesis of depressive affect. *Psychiatry* 74, 5–13. doi: 10.1521/psyc.2011.74.1.5

Peschard, V., Philippot, P., Joassin, F., and Rossignol, M. (2013). The impact of the stimulus features and task instructions on facial processing in social anxiety: an ERP investigation. *Biol. Psychol.* 93, 88–96. doi: 10.1016/j.biopsycho.2013.01.009

Phan, K. L., Fitzgerald, D. A., Nathan, P. J., and Tancer, M. E. (2006). Association between amygdala hyperactivity to harsh faces and severity of social anxiety in generalized social phobia. *Biol. Psychiatry* 59, 424–429. doi: 10.1016/j.biopsych.2005.08.012

Pitterman, H., and Nowicki, S. (2004). A test of the ability to identify emotion in human standing and sitting postures: the diagnostic analysis of nonverbal accuracy-2 posture test (DANVA2-POS). *Genet. Soc. Gen. Psychol. Monogr.* 130, 146–162. doi: 10.3200/mono.130.2.146-162

Pönkänen, L. M., and Hietanen, J. K. (2012). Eye contact with neutral and smiling faces: effects on autonomic responses and frontal EEG asymmetry. *Front. Hum. Neurosci.* 6:122. doi: 10.3389/fnhum.2012.00122

Poulin, M. J., Holman, E. A., and Buffone, A. (2012). The neurogenetics of nice: receptor genes for oxytocin and vasopressin interact with threat to predict prosocial behavior. *Psychol. Sci.* 23, 446–452. doi: 10.1177/0956797611428471

Quadflieg, S., Mohr, A., Mentzel, H.-J., Miltner, W. H. R., and Straube, T. (2008). Modulation of the neural network involved in the processing of anger prosody: the role of task-relevance and social phobia. *Biol. Psychol.* 78, 129–137. doi: 10.1016/j.biopsycho.2008.01.014

Quadflieg, S., Wendt, B., Mohr, A., Miltner, W. H. R., and Straube, T. (2007). Recognition and evaluation of emotional prosody in individuals with generalized social phobia: a pilot study. *Behav. Res. Ther.* 45, 3096–3103. doi: 10.1016/j.brat.2007.08.003

Quirin, M., Düsing, R., and Kuhl, J. (2013). Implicit affiliation motive predicts correct intuitive judgment. *J. Individ. Differences* 34, 24–31. doi: 10.1027/1614-0001/a000086

Rapee, R. M., and Heimberg, R. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behav. Res. Ther.* 35, 741–756. doi: 10.1016/s0005-7967(97)00022-3

Rapee, R. M., and Lim, L. (1992). Discrepancy between self- and observer ratings of performance in social phobics. *J. Abnorm. Psychol.* 101, 728–731. doi: 10.1037//0021-843x.101.4.728

Rapee, R. M., McCallum, S. L., Melville, L. F., Ravenscroft, H., and Rodney, J. M. (1994). Memory bias in social phobia. *Behav. Res. Ther.* 32, 89–99. doi: 10.1016/0005-7967(94)90087-6

Reed, C. L., Stone, V. E., and McGoldrick, J. E. (2006). "Not just posturing: configural processing of the human body," in *Human Body Perception From the Inside Out: Advances in Visual Cognition*, eds G. Knoblich, I. M. Thornton, M. Grosjean and M. Shiffrar (New York, NY, US: Oxford University Press), 229–258.

Reed, C. L., Stone, V. E., Bozova, S., and Tanaka, J. (2003). The body-inversion effect. *Psychol. Sci.* 14, 302–308. doi: 10.1111/1467-9280.14431

Riem, M. M. E., Bakermans-Kranenburg, M. J., Huffmeijer, R., and van Ijzendoorn, M. H. (2013). Does intranasal oxytocin promote prosocial behavior to an excluded fellow player? A randomized-controlled trial with Cyberball. *Psychoneuroendocrinology* 38, 1418–1425. doi: 10.1016/j.psyneuen.2012.12.023

Rinck, M., and Becker, E. S. (2005). A comparison of attentional biases and memory biases in women with social phobia and major depression. *J. Abnorm. Psychol.* 114, 62–74. doi: 10.1037/0021-843x.114.1.62

Rodebaugh, T. L. (2009). Social phobia and perceived friendship quality. *J. Anxiety Disord.* 23, 872–878. doi: 10.1016/j.janxdis.2009.05.001

Roelofs, K., Putman, P., Schouten, S., Lange, W.-G., Volman, I., and Rinck, M. (2010). Gaze direction differentially affects avoidance tendencies to happy and angry faces in socially anxious individuals. *Behav. Res. Ther.* 48, 290–294. doi: 10.1016/j.brat.2009.11.008

Rowell, T. E. (1974). The concept of social dominance. *Behav. Biol.* 11, 131–154. doi: 10.1016/s0091-6773(74)90289-2

Rule, N. O., Adams, G. R., Ambady, N., and Freeman, J. B. (2012). Perceptions of dominance following glimpses of faces and bodies. *Perception* 41, 687–706. doi: 10.1068/p7023

Safren, S. A., Heimberg, R. G., Horner, K. J., Juster, H. R., Schneier, F. R., and Liebowitz, M. R. (1999). Factor structure of social fears: the liebowitz social anxiety scale. *J. Anxiety Disord.* 13, 253–270. doi: 10.1016/s0887-6185(99)00003-1

Sapolsky, R. M. (2005). The influence of social hierarchy on primate health. *Science* 308, 648–652. doi: 10.1126/science.1106477

Sauter, D. A., and Eimer, M. (2010). Rapid detection of emotion from human vocalizations. *J. Cogn. Neurosci.* 22, 474–481. doi: 10.1162/jocn.2009.21215

Scherer, K. R. (1981). "Vocal indicators of stress," in *Speech Evaluation in Psychiatry*, ed J. K. Darby (New York, NY: Grune & Stratton), 171–187.

Schmitz, J., Scheel, C. N., Rigon, A., Gross, J. J., and Blechert, J. (2012). You don't like me, do you? Enhanced ERP responses to averted eye gaze in social anxiety. *Biol. Psychol.* 91, 263–269. doi: 10.1016/j.biopsycho.2012.07.004

Schneier, F. R., Kent, J. M., Star, A., and Hirsch, J. (2009). Neural circuitry of submissive behavior in social anxiety disorder: a preliminary study of response to direct eye gaze. *Psychiatry Res.* 173, 248–250. doi: 10.1016/j.pscychresns.2008.06.004

Schneier, F. R., Rodebaugh, T. L., Blanco, C., Lewin, H., and Liebowitz, M. R. (2011). Fear and avoidance of eye contact in social anxiety disorder. *Compr. Psychiatry* 52, 81–87. doi: 10.1016/j.comppsych.2010.04.006

Schofield, C. A., Johnson, A. L., Inhoff, A. W., and Coles, M. E. (2012). Social anxiety and difficulty disengaging threat: evidence from eye-tracking. *Cogn. Emot.* 26, 300–311. doi: 10.1080/02699931.2011.602050

Schug, J., Matsumoto, D., Horita, Y., Yamagishi, T., and Bonnet, K. (2010). Emotional expressivity as a signal of cooperation. *Evol. Hum. Behav.* 31, 87–94. doi: 10.1016/j.evolhumbehav.2009.09.006

Schultheiss, O. C., and Wirth, M. M. (2008). "Biopsychological aspects of motivation," in *Motivation and Action*, 2nd Edn., ed J. H. H. Heckhausen (New York, NY, US: Cambridge University Press), 247–271.

Sellers, J. G., Mehl, M. R., and Josephs, R. A. (2007). Hormones and personality: testosterone as a marker of individual differences. *J. Res. Pers.* 41, 126–138. doi: 10.1016/j.jrp.2006.02.004

Shariff, A. F., and Tracy, J. L. (2009). Knowing who's boss: implicit perceptions of status from the nonverbal expression of pride. *Emotion* 9, 631–639. doi: 10.1037/a0017089

Shaver, P. R., Segev, M., and Mikulincer, M. (2011). "A behavioral systems perspective on power and aggression," in *Human aggression and Violence: Causes, Manifestations and Consequences*, eds P. R. Shaver and M. Mikulincer (Washington, DC, US: American Psychological Association), 71–87.

Siegman, A. W. (1987). "The telltale voice: nonverbal messages of verbal communication," in *Nonverbal Behavior and Communication*, eds A. W. Siegman and S. Feldstein (NJ: Hillsdale), 351–433.

Silk, J. B. (2007). Social components of fitness in primate groups. *Science* 317, 1347–1351. doi: 10.1126/science.1140734

Sloman, K., and Armstrong, J. D. (2002). Physiological effects of dominance hierarchies: laboratory artefacts or natural phenomena? *J. Fish Biol.* 61, 1–23. doi: 10.1111/j.1095-8649.2002.tb01733.x

Spokas, M., Luterek, J. A., and Heimberg, R. G. (2009). Social anxiety and emotional suppression: the mediating role of beliefs. *J. Behav. Ther. Exp. Psychiatry* 40, 283–291. doi: 10.1016/j.jbtep.2008.12.004

Stanton, S. J., Hall, J. L., and Schultheiss, O. C. (2010). "Properties of motive-specific incentives," in *Implicit Motives*, eds O. C. Schultheiss and J. C. Brunstein (New York: Oxford University Press), 245–278.

Stanton, S. J., and Schultheiss, O. C. (2007). Basal and dynamic relationships between implicit power motivation and estradiol in women. *Horm. Behav.* 52, 571–580. doi: 10.1016/j.yhbeh.2007.07.002

Staugaard, S. R. (2010). Threatening faces and social anxiety: a literature review. *Clin. Psychol. Rev.* 30, 669–690. doi: 10.1016/j.cpr.2010.05.001

Stein, M. B. (2002). Increased amygdala activation to angry and contemptuous faces in generalized social phobia. *Arch. Gen. Psychiatry* 59, 1027–1034. doi: 10.1001/archpsyc.59.11.1027

Stein, M. B., Goldin, P. R., Sareen, J., Zorrilla, L. T. E., and Brown, G. G. (2002). Increased amygdala activation to angry and contemptuous faces in generalized social phobia. *Arch. Gen. Psychiatry* 59, 1027–1034. doi: 10.1001/archpsyc.59.11.1027

Stein, D. J., Kasper, S., Andersen, E. W., Nil, R., and Lader, M. (2004). Escitalopram in the treatment of social anxiety disorder: analysis of efficacy for different clinical subgroups and symptom dimensions. *Depress. Anxiety* 20, 175–181. doi: 10.1002/da.20043

Stein, M. B., and Kean, Y. M. (2000). Disability and quality of life in social phobia: epidemiologic findings. *Am. J. Psychiatry* 157, 1606–1613. doi: 10.1176/appi.ajp.157.10.1606

Stel, M., Dijk, W. W., Smith, P. K., and Djalal, F. M. (2012). Lowering the pitch of your voice makes you feel more powerful and think more abstractly. *Soc. Psychol. Personal. Sci.* 3, 497–502. doi: 10.1177/1948550611427610

Stel, M., van Baaren, R. B., Blascovich, J., van Dijk, E., McCall, C., Pollmann, M. M. H., et al. (2010). Effects of a priori liking on the elicitation of mimicry. *Exp. Psychol.* 57, 412–418. doi: 10.1027/1618-3169/a000050

Stel, M., and van Knippenberg, A. (2008). The role of facial mimicry in the recognition of affect. *Psychol. Sci.* 19, 984–985. doi: 10.1111/j.1467-9280.2008.02188.x

Stevens, S., Hofmann, M., Kiko, S., Mall, A. K., Steil, R., Bohus, M., et al. (2010). What determines observer-rated social performance in individuals with social anxiety disorder? *J. Anxiety Disord.* 24, 830–836. doi: 10.1016/j.janxdis.2010.06.005

Stienen, B. M. C., and de Gelder, B. (2011). Fear detection and visual awareness in perceiving bodily expressions. *Emotion* 11, 1182–1189. doi: 10.1037/a0024032

Straube, T., Mentzel, H.-J., and Miltner, W. H. R. (2005). Common and distinct brain activation to threat and safety signals in social phobia. *Neuropsychobiology* 52, 163–168. doi: 10.1159/000087987

Taylor, S. E. (2006). Tend and befriend: biobehavioral bases of affiliation under stress. *Curr. Dir. Psychol. Sci.* 15, 273–277. doi: 10.1111/j.1467-8721.2006.00451.x

Taylor, C. T., and Alden, L. E. (2011). To see ourselves as others see us: an experimental integration of the intra and interpersonal consequences of self-protection in social anxiety disorder. *J. Abnorm. Psychol.* 120, 129–141. doi: 10.1037/a0022127

Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A. R., and Updegraff, J. A. (2000). Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight. *Psychol. Rev.* 107, 411–429. doi: 10.1037/0033-295x.107.3.411

Terburg, D., Aarts, H., and van Honk, J. (2012). Memory and attention for social threat: anxious hypercoding-avoidance and submissive gaze aversion. *Emotion* 12, 666–672. doi: 10.1037/a0027201

Terburg, D., and van Honk, J. (2013). Approach-avoidance versus dominance-submissiveness: a multilevel neural framework on how testosterone promotes social status. *Emot. Rev.* 5, 296–302. doi: 10.1177/1754073913477510

Thomsen, L., Frankenhuis, W. E., Ingold-Smith, M., and Carey, S. (2011). Big and mighty: preverbal infants mentally represent social dominance. *Science* 331, 477–480. doi: 10.1126/science.1199198

Tiedens, L. Z., and Fragale, A. R. (2003). Power moves: complementarity in dominant and submissive nonverbal behavior. *J. Pers. Soc. Psychol.* 84, 558–568. doi: 10.1037/0022-3514.84.3.558

Tinlin, R. M., Watkins, C. D., Welling, L. L. M., DeBruine, L. M., Al-Dujaili, E. A. S., and Jones, B. C. (2013). Perceived facial adiposity conveys information about women's health. *Br. J. Psychol.* 104, 235–248. doi: 10.1111/j.2044-8295.2012.02117.x

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., and Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion* 13, 724–738. doi: 10.1037/a0032335

Todorov, A., Pakrashi, M., and Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Soc. Cogn.* 27, 813–833. doi: 10.1521/soco.2009.27.6.813

Tracy, J. L., and Robins, R. W. (2004). Show your pride: evidence for a discrete emotion expression. *Psychol. Sci.* 15, 194–197. doi: 10.1111/j.0956-7976.2004.01503008.x

Tracy, J. L., and Robins, R. W. (2007). The prototypical pride expression: development of a nonverbal behavior coding system. *Emotion* 7, 789–801. doi: 10.1037/1528-3542.7.4.789

Tracy, J. L., Shariff, A. F., Zhao, W., and Henrich, J. (2013). Cross-cultural evidence that the nonverbal expression of pride is an automatic status signal. *J. Exp. Psychol. Gen.* 142, 163–180. doi: 10.1037/a0028412

Van den Stock, J., Righart, R., and de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion* 7, 487–494. doi: 10.1037/1528-3542.7.3.487

Veit, R., Flor, H., Erb, M., Hermann, C., Lotze, M., Grodd, W., et al. (2002). Brain circuits involved in emotional learning in antisocial behavior and social phobia in humans. *Neurosci. Lett.* 328, 233–236. doi: 10.1016/s0304-3940(02)00519-0

Voncken, M. J., and Bögels, S. M. (2008). Social performance deficits in social anxiety disorder: reality during conversation and biased perception during speech. *J. Anxiety Disord.* 22, 1384–1392. doi: 10.1016/j.janxdis.2008.02.001

Vrana, S. R., and Gross, D. (2004). Reactions to facial expressions: effects of social context and speech anxiety on responses to neutral, anger and joy expressions. *Biol. Psychol.* 66, 63–78. doi: 10.1016/j.biopsycho.2003.07.004

Vrijsen, J. N., Lange, W.-G., Becker, E. S., and Rinck, M. (2010). Socially anxious individuals lack unintentional mimicry. *Behav. Res. Ther.* 48, 561–564. doi: 10.1016/j.brat.2010.02.004

Walker, A. S., Nowicki, S., Jones, J., and Heimann, L. (2011). Errors in identifying and expressing emotion in facial expressions, voices and postures unique to social anxiety. *J. Genet. Psychol.* 172, 293–301. doi: 10.1080/00221325.2010.535224

Wallace, S. T., and Alden, L. E. (1997). Social phobia and positive social events: the price of success. *J. Abnorm. Psychol.* 106, 416–424. doi: 10.1037/0021-843x.106.3.416

Weeks, J., Heimberg, R. G., and Heuer, R. (2011). Exploring the role of behavioral submissiveness in social anxiety. *Soc. Clin. Psychol.* 30, 217–249. doi: 10.1521/jscp.2011.30.3.217

Weeks, J. W., Lee, C.-Y., Reilly, A. R., Howell, A. N., France, C., Kowalsky, J. M., et al. (2012). "The sound of fear": assessing vocal fundamental frequency as a physiological indicator of social anxiety disorder. *J. Anxiety Disord.* 26, 811–822. doi: 10.1016/j.janxdis.2012.07.005

Weisfeld, G. E., and Beresford, J. M. (1982). Erectness of posture as an indicator of dominance or success in humans. *Motiv. Emot.* 6, 113–131. doi: 10.1007/bf00992459

Weisman, O., Aderka, I. M., Marom, S., Hermesh, H., and Gilboa-Schechtman, E. (2011). Social rank and affiliation in social anxiety disorder. *Behav. Res. Ther.* 49, 399–405. doi: 10.1016/j.brat.2011.03.010

West-Eberhard, M. J. (1979). Sexual selection, social competition and evolution. *Proc. Am. Philos. Soc.* 123, 222–234.

Wieser, M. J., McTeague, L. M., and Keil, A. (2011). Sustained preferential processing of social threat cues: bias without competition? *J. Cogn. Neurosci.* 23, 1973–1986. doi: 10.1162/jocn.2010.21566

Wieser, M. J., Pauli, P., Alpers, G. W., and Mühlberger, A. (2009). Is eye to eye contact really threatening and avoided in social anxiety?–An eye-tracking and psychophysiology study. *J. Anxiety Disord.* 23, 93–103. doi: 10.1016/j.janxdis.2008.04.004

Wieser, M. J., Pauli, P., Grosseibl, M., Molzow, I., and Mühlberger, A. (2010). Virtual social interactions in social anxiety–the impact of sex, gaze and interpersonal distance. *Cyberpsychol. Behav. Soc. Netw.* 13, 547–554. doi: 10.1089/cyber.2009.0432

Wirth, J. H., Sacco, D. F., Hugenberg, K., and Williams, K. D. (2010). Eye gaze as relational evaluation: averted eye gaze leads to feelings of ostracism and relational devaluation. *Pers. Soc. Psychol. Bull.* 36, 869–882. doi: 10.1177/0146167210370032

Wolff, S. E., and Puts, D. A. (2010). Vocal masculinity is a robust dominance signal in men. *Behav. Ecol. Sociobiol.* 64, 1673–1683. doi: 10.1007/s00265-010-0981-5

Yoon, K. L., Fitzgerald, D. A., Angstadt, M., McCarron, R. A., and Phan, K. L. (2007). Amygdala reactivity to emotional faces at high and low intensity in generalized social phobia: a 4-Tesla functional MRI study. *Psychiatry Res.* 154, 93–98. doi: 10.1016/j.pscychresns.2006.05.004

Young, T. J., and French, L. A. (1998). Heights of U.S. Presidents: a trend analysis for 1948-1996. *Percept. Mot. Skills.* 87, 321–322. doi: 10.2466/pms.1998.87.1.321

Zitek, E. M., and Tiedens, L. Z. (2012). The fluency of social hierarchy: the ease with which hierarchical relationships are seen, remembered, learned and liked. *J. Pers. Soc. Psychol.* 102, 98–115. doi: 10.1037/a0025345

Zuckerman, M., and Driver, R. E. (1989). What sounds beautiful is good: the vocal attractiveness stereotype. *J. Nonverbal Behav.* 13, 67–82. doi: 10.1007/bf00990791

Zuroff, D. C., Fournier, M. A., Patall, E. A., and Leybman, M. J. (2010). Steps toward an evolutionary personality psychology: individual differences in the social rank domain. *Can. Psychol.* 51, 58–66. doi: 10.1037/a0018472

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 August 2013; accepted: 10 December 2013; published online: 31 December 2013.

Citation: Gilboa-Schechtman E and Shachar-Lavie I (2013) More than a face: a unified theoretical perspective on nonverbal social cue processing in social anxiety. Front. Hum. Neurosci. 7:904. doi: 10.3389/fnhum.2013.00904

This article was submitted to the journal Frontiers in Human Neuroscience.

Copyright © 2013 Gilboa-Schechtman and Shachar-Lavie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Molecular Psychiatry
https://doi.org/10.1038/s41380-021-01171-5

EXPERT REVIEW

Psychobiological risk factors for suicidal thoughts and behaviors in adolescence: a consideration of the role of puberty

Tiffany C. Ho[1,2] · Anthony J. Gifuni[1,3] · Ian H. Gotlib[1]

Received: 17 April 2020 / Revised: 3 May 2021 / Accepted: 11 May 2021
© The Author(s) 2021. This article is published with open access

Abstract

Suicide is the second leading cause of death among adolescents. While clinicians and researchers have begun to recognize the importance of considering multidimensional factors in understanding risk for suicidal thoughts and behaviors (STBs) during this developmental period, the role of puberty has been largely ignored. In this review, we contend that the hormonal events that occur during puberty have significant effects on the organization and development of brain systems implicated in the regulation of social stressors, including amygdala, hippocampus, striatum, medial prefrontal cortex, orbitofrontal cortex, and anterior cingulate cortex. Guided by previous experimental work in adults, we also propose that the influence of pubertal hormones and social stressors on neural systems related to risk for STBs is especially critical to consider in adolescents with a neurobiological sensitivity to hormonal changes. Furthermore, facets of the pubertal transition, such as pubertal timing, warrant deeper investigation and may help us gain a more comprehensive understanding of sex differences in the neurobiological and psychosocial mechanisms underlying adolescent STBs. Ultimately, advancing our understanding of the pubertal processes that contribute to suicide risk will improve early detection and facilitate the development of more effective, sex-specific, psychiatric interventions for adolescents.

Introduction

Suicide is the second leading cause of death among adolescents ages 10–24 years worldwide [1] and accounts for 17.7% of all deaths in youth ages 15–24 years in the United States [2]. Although suicide has a relatively low prevalence, over the past 10 years suicide rates in the US have increased by 76% in youth ages 15–19 years and, more alarmingly, by 300% in youth ages 10–14 years [3]. Despite the urgency of suicide as a public health problem, research examining predictors of suicidal thoughts and behaviors (STBs) has had limited clinical impact, particularly in adolescents [4, 5]. Adolescents report a higher proportion of suicidal ideation and attempts than do adults, making this a critical developmental period during which to identify early risk factors for STBs [6]. Indeed, the fact that adolescent samples represent only 20% of the literature on predictors of STBs further underscores the need for more research with this age group [5].

While clinicians and researchers recognize that predictors of STBs are multidimensional—including genetic, epigenetic, neurobiological, psychosocial, and environmental factors—the marked increase in the prevalence of STBs (and mental health conditions more generally) during adolescence [7] suggests that this developmental period, in which the confluence of these multidimensional influences contributes to elevated risk, is unique [8, 9]. In this context, we contend that one critical factor is puberty. The onset of puberty initiates a neuroendocrine cascade that shapes the maturation of neural circuits that underlie a range of socioemotional and cognitive functioning, including emotion regulation and impulse control. Impaired emotion regulation and impulse control both significantly contribute to the emergence of mental disorders in the face of environmental risk factors (e.g., life stress) that influence neurodevelopmental trajectories, result in the generation and experience of additional psychosocial stress, and, thus, may represent an important pathway to developing

✉ Tiffany C. Ho
tiffany.ho@ucsf.edu

[1] Department of Psychology, Stanford University, Stanford, CA, USA

[2] Department of Psychiatry and Weill Institute for Neuroscience, University of California, San Francisco, San Francisco, CA, USA

[3] Psychiatry Department and Douglas Mental Health University Institute, McGill University, Montréal, QC, Canada

Published online: 11 June 2021

STBs. Other notable aspects of development, such as pubertal timing, and factors that may drive atypical pubertal timing (e.g., exposure to early adversity), are also recognized as important in affecting the development of adolescent-onset psychiatric disorders, including STBs [10]. Indeed, atypical pubertal timing results not only in corresponding changes in neuroendocrine systems but also significant shifts in social identity and perceived status may explain sex differences in the emergence and course of STBs (and related mental health conditions that substantially increase the risk of STBs).

The goal of this review paper is to elucidate how pubertal changes contribute to STBs in adolescence. While we do not regard puberty-related processes as primary determinants of STBs, we posit that exposure to sex hormones during adolescence initiates a period of plasticity in neural circuits that are sensitive to social context (including social stressors that amplify emotion dysregulation and impulse control in ways that increase risk for STBs) and that may be target mechanisms for treatment of STBs, We further contend that these processes are especially critical to consider in adolescents who have existing vulnerabilities, including neurobiological sensitivity to hormonal changes, exposure to adverse psychosocial experiences during early development, and underlying mental disorders. In the following sections, we: (1) present an overarching framework and highlight specific brain circuits involved in social cognition, emotion regulation, and impulse control that are relevant to understanding STBs in adolescence; (2) review basic experimental findings in adults that show the effect of pubertal hormones on behaviors relevant to suicide risk and survey the conflicting correlational literature on hormones with STBs in adolescents; (3) discuss the role of pubertal timing and sex differences in these processes; and (4) advance recommendations for future research in this area.

Puberty as a driver of neuroendocrine mechanisms relevant for understanding adolescent STBs

Contemporary theories of suicide have been informed primarily by data from adults [11, 12], it is unclear whether these theories—or specific components of associated models—extend to adolescents. Moreover, few theories of suicidality explicitly integrate biological factors, including endocrine or neural factors (with exceptions [12–14]). Here, we posit that adolescent STBs result from a pathological response to stress during a time when the neurobiological systems that regulate stress are recalibrating. There is extensive evidence that STBs in adolescents are often preceded by life stressors, particularly stressors characterized by interpersonal rejection, loss, or conflict [15–17]. Social rejection in particular is a commonly experienced stressor during adolescence due to unstable romantic and peer relationships, and is a potent trigger for negative emotions [18, 19]. In this context, emotion-related impulsivity—the tendency to react impulsively during experiences of heighted affective states [20]—may contribute to adolescent risk for STBs. Indeed, both peer-related stressors [21, 22] and emotion-related impulsivity [23] have been shown separately to predict STBs in adolescents.

We argue here that hormonal changes during puberty alter the development of brain circuits implicated in the regulation of social stressors. The brain is a target organ for all of the pubertal hormones we review, and receptors for these hormones are expressed abundantly in several key structures that comprise brain networks governing social cognition, emotion regulation, and impulse control generally (and social rejection and emotion-related impulsivity specifically) [24, 25], including the amygdala, hippocampus, striatum, anterior cingulate cortex (ACC), medial prefrontal cortex (mPFC), and orbitofrontal cortex (OFC). The amygdala, hippocampus, and ACC comprise a network of regions implicated in detecting social salience; the amygdala and mPFC comprise a key emotion regulatory circuit; the striatum and frontal regions including OFC comprise circuits underlying reward valuation and impulse control. It is important to note that many of these regions already exhibit sexual dimorphism during adolescence, including larger amygdala and hippocampal volumes in boys, greater variance in the hippocampus and striatum in boys than in girls, sex differences in white matter organization of callosal, cerebellar, and long-range association tracts (for a review, see [26]), and sex differences in the developmental trajectories of frontoparietal networks [27]. Moreover, several of these regions have been highlighted in a recent review of neuroimaging markers associated with STBs across the lifespan [28]; however, those authors did not consider the importance of pubertal development in explaining the neuroendocrine basis of the emergence of STBs. We contend that this is a critical factor to consider, given emerging evidence that sex differences in these brain circuits during adolescence appear to be explained, at least in part, by changes in pubertal hormones [29, 30]. Thus, we contend that the social landscape, and most notably environmental stressors, experienced by boys and girls begins to differ during puberty, and that these differences, in conjunction with sex-specific hormonal effects on brain maturation, may explain important sex differences in STBs that emerge by mid-adolescence.

Pubertal hormones and suicide-relevant thoughts and behaviors

Puberty is composed of two phases: *adrenarche* and *gonadarche*. During adrenarche, which typically begins around ages 7–8 years, the adrenal glands produce increasing levels

of the hormones dehydroepiandrosterone (DHEA) and testosterone [31, 32]. Gonadarche is a longer process that typically takes 4–5 years, beginning around ages 9–10 and occurring on average a year earlier in girls than in boys [33]. Gonadarche is triggered by the activation of hypothalamic-pituitary-gonadal (HPG) axis, which leads to rising levels of luteinizing hormone (LH) and follicle-stimulating hormone (FSH). The release of LH and FSH initiates the development of the gonads, which, in turn, leads to increases in sex hormones—specifically, testosterone in boys and estradiol (the predominant estrogen during adolescence) and progesterone in girls—and, by early to mid-adolescence, the development of secondary sex characteristics and other physical changes [32]. Importantly, these hormones cross the blood brain barrier, influence brain development, and affect a wide variety of signaling pathways (e.g., neurotransmitter activity) that underlie mood and cognition [25]. Thus, puberty involves transformation across virtually every psychobiological domain—endocrine, neural, physical, cognitive, and socioemotional—and represents a vulnerable time during which STBs may emerge.

Prevalence rates of STBs begin to increase after age 12 but peak in later adolescence, suggesting that it is not simply the rise in these hormones that accounts for the increase in STBs, given that these endocrine changes begin much earlier than age 12. Instead, the organizational effects of pubertal hormones play a major role in risk for STBs in adolescence. Organizational effects are those that occur during sensitive periods of development that lay the foundation for sex-typical brain and behavioral phenotypes and, thus, have an impact even in the absence of circulating levels of the hormones. In contrast, activational effects facilitate the expression of behaviors under specific contexts but are temporary and occur only when the hormones in question are present [34]. Evidence for organizational effects of hormones during adolescence comes from studies demonstrating that the same developmental processes (e.g., neurogenesis, synaptic pruning, dendritic branching, apoptosis) that occur during the perinatal period as a result of surging levels of gonadal hormones also occur during puberty [35, 36]. These studies suggest that puberty opens a sensitive window for experience-dependent plasticity in neural circuits that underlie higher-order processing of social stimuli, thereby rendering adolescence—a time of increased exposure to social stressors—a vulnerable period for the onset of stress-related mental health problems (Fig. 1). At the same time, it is possible that ongoing fluctuations in hormone levels are also related to STB risk through their activational effects on neurotransmitter systems, particularly in the context of responding to social stressors, and that these effects are more relevant for adolescents with a neurobiological sensitivity to hormonal changes.

Below, we review studies investigating ovarian (estradiol and progesterone), testicular (testosterone), and adrenal (DHEA) hormones in relation to STBs and, importantly, their impact on the structural development and function of specific brain circuits implicated in social cognition, emotion regulation, and impulse control. A summary of this section is outlined in Table 1 and Fig. 1B. Before reviewing the correlational literature in adolescents, within each respective section, we first review basic experimental studies that provide insight concerning the effects of estradiol, progesterone, testosterone, and DHEA on human behaviors relevant to suicide risk in adults in order to provide a more reliable context for the actions of these hormones.

Finally, because ovarian hormones have been investigated almost exclusively in females in this context and, similarly, androgen hormones overwhelmingly in males, our review of the association of these hormones with brain and behavioral outcomes inherently raises issues of sex differences or sex-specific effects in these processes; we discuss these issues in more detail in a later section (D. Consideration of sex differences).

Ovarian hormones (estradiol and progesterone)

Estradiol, progesterone, and their neurosteroid metabolites all increase and begin to fluctuate cyclically in girls during puberty [34]. Consequently, the vast majority of research in this area has focused on females (for reviews, see [14, 37]). To provide a context for understanding the effects of estradiol and progesterone administration on behaviors relevant to suicide risk, we will first briefly review studies that controlled hormone conditions in women with affective symptoms and premenstrual dysphoric disorder (PMDD), for whom suicide risk is elevated compared to the general population [38, 39].

Decades of research have shown that there are no differences between women with and without PMDD in ovarian hormone levels or related neurosteroids (e.g., allopregnanolone) but that suppressing ovarian hormones reduces or eliminates symptoms of PMDD (for a review, see [40]). A recent study that controlled for ovarian hormone secretion and exposure in women with PMDD has helped to clarify these two seemingly opposing findings [41]. In that study, women with PMDD who responded to a gonadotropin-releasing hormone agonist treatment were given placebo for one month before being administered continuous estradiol/progesterone for three months; researchers found that changes from low to high levels of ovarian hormones, but not absolute levels of ovarian hormones, were associated with increases in negative affect [41]. Together, these data suggest that neurobiological sensitivity to hormone *changes* is an important factor that may explain certain clinical phenomena, such as PMDD and suicide risk. Indeed, a recent review has covered this topic extensively in adult women, demonstrating that cyclical

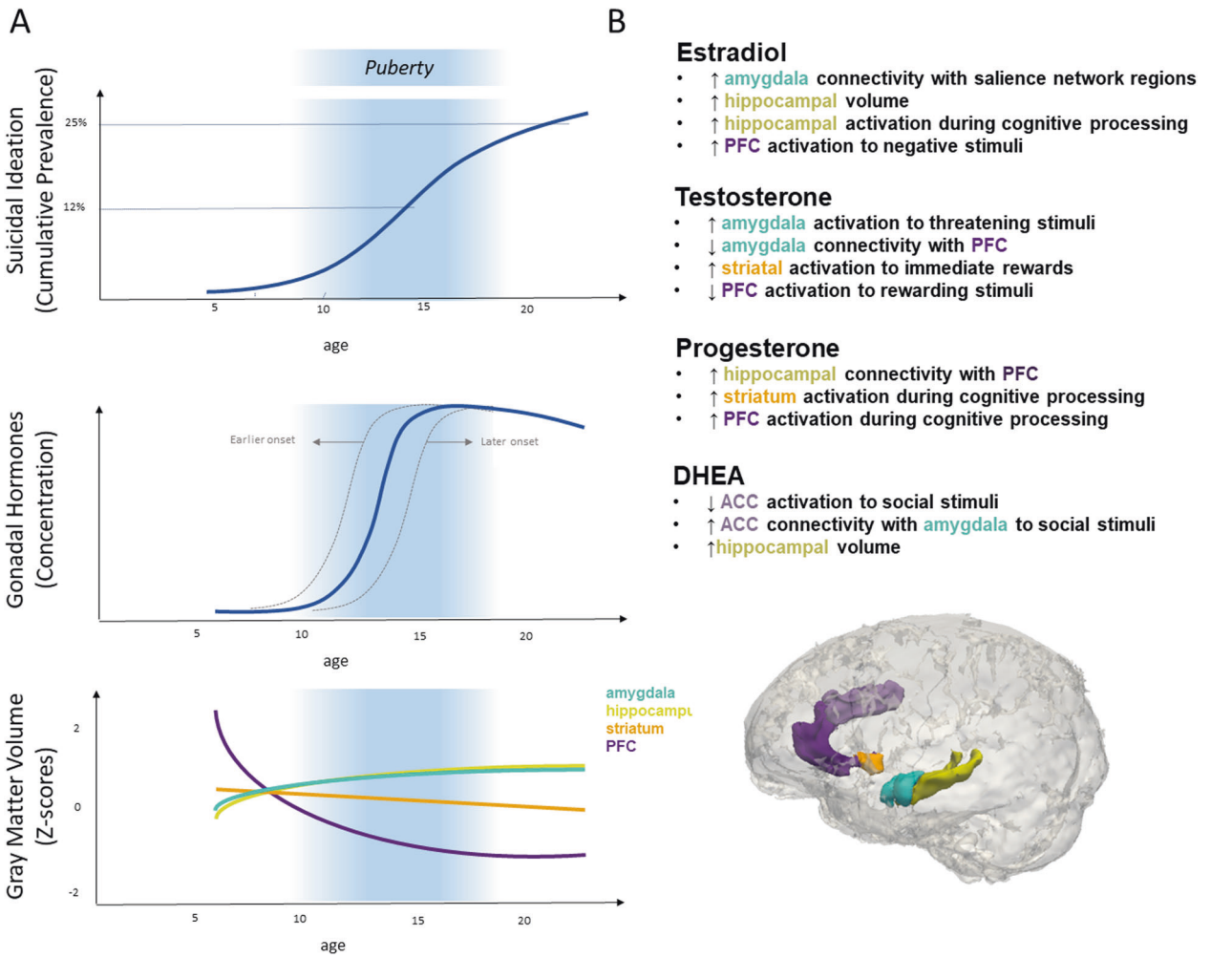


Fig. 1 Summary of prevalence rates of suicidal ideation, concentration of sex steroids, and brain volume as a function of age and typical associations between pubertal hormones and brain structures reported the extant literature. A) Graphical depictions of prevalence rates of suicidal ideation, concentrations of sex steroids, and brain volume as a function of age. Shaded region indicates puberty. The schematized trajectories of gray matter volume adjusted for total brain volume are based on data reported in [162]. **B)** Summary of typical associations between pubertal hormones and brain structures from both adolescent and adult samples. ACC anterior cingulate cortex, AMYG amygdala, HPC hippocampus, MPFC medial prefrontal cortex, OFC orbitofrontal cortex, STM striatum.

hormone changes may play an important role in "acute risk for daily suicidal ideation, planning, and intent" in individuals with sensitivity to hormone changes [14]. We have extended this theoretical model to adolescents, proposing that in those who possess a neurobiological sensitivity to hormonal changes, the normative fluctuations during this transitional period, coupled with adolescent-typical experiences of greater exposure to life stressors, may exacerbate the effects of these processes on emotion regulatory and impulse control circuits. Indeed, results from studies that have investigated suicide risk in adult women across the menstrual cycle have been inconsistent. These studies have typically assessed levels of estradiol and progesterone during different phases of the menstrual cycle [42–44], and in women with low levels of estradiol and progesterone (e.g., amenorrhea or menopause [45]). While some investigators have reported that a higher risk of suicide attempts and more severe suicidal thoughts and intentions are associated with relatively low or declining levels of estradiol and progesterone (i.e., during the early follicular/menstrual or pre-menstrual phases), other researchers have not found differences in estradiol or progesterone between depressed women with and without STBs [46] or effects of the menstrual cycle on suicide attempts in women with PMDD [39].

The fact that there has been no evidence of differences in ovarian hormone levels or in cyclical changes in female adolescents with and without STBs is consistent with the previously aforementioned studies in adult women with reproductive mood disorders where the effect of menstrual cycle and/or ovarian hormone levels on STBs in adult women is absent. One interpretation is that it is the

Table 1 Summary of studies examining gonadal and adrenal hormones in relation to suicide attempts and suicidal ideation.

Publication	Sample size and characteristics	Age (years)	Sex	Psychiatric condition	Suicide-related outcome	Study design and methods	Findings	Additional notes
Afzali et al. 2012	81 suicide attempters	Mean = 23.63 SD = 8.41 Range = 15–55	F	Assorted (25 Past mental disorder, 22 Previous suicide attempt)	History of Suicide Attempts	Structured interview over 6 months after attempt	Suicide attempts were not associated with menstrual cycle phase.	Patients with irregular menstrual cycles were excluded.
Baca-Garcia et al. 2010[a]	281 suicide attempters 176 healthy controls	Mean = 30.8 SD = 8.8 Range = 18–92	F	Assorted (229 Mood disorder, 229 SUD, 275 Previous psychiatric treatment)	Recent Suicide Attempts and Recent Suicidal Ideation	Blood sample within 24 h of attempt: estradiol, progesterone, LH, FSH	Suicide attempts were was more likely during the follicular phase. Suicide intent severity was elevated during low-estrogen/low-progesterone states (pre-menstrual phase, amenorreha, menopause)	
Butterfield et al. 2005	130 inpatients	Mean = 49.4 SD = 8.13	M	PTSD	Recent History of Suicide Attempts (past 6 months) and Suicidal Ideation	Blood: DHEA, androstenedione, testosterone, estradiol	Suicide attempters had higher DHEA than nonattempters	
Cayköylü et al. 2004[a]	52 suicide attempters 50 healthy controls	Mean = 26.51 SD = 7.82 Range = Not Reported	F	Assorted (8 PMDD, 1 SCZ, 2 MDD, 1 OCD)	Recent Suicide Attempts	Blood sample within 12 h of attempt: estradiol, progesterone Menstrual status determined with self-report.	Suicide attempts were more frequent during the follicular phase. Estradiol and progesterone levels were not different in suicide attempters compared to healthy controls.	Patients attempting suicide with OD or admitted to the ICU were excluded.
Chatzittofis et al. 2013	28 suicide attempters (10 female, 18 male) 19 healthy controls (7 female, 12 male)	SA: Mean = 44 SD = 14.6 Range = 26–66 HC: Mean = 30 SD = Not Reported range = 23–48	Both	Assorted (14 Mood disorder, 4 Anxiety Disorder, 9 SUD, 19 PD)	History of Suicide Attempts	CSF: DHEA-S, DHEA, cortisol, and 5-HIAA	In males, suicide attempters had higher CSF DHEA-S levels compared to healthy controls. In females, no significant differences.	Exposure to early adversity (e.g., interpersonal violence) correlated negatively with cortisol/DHEA-S ratio
Dogra et al. 2007[a]	217 suicide decedents 237 non-suicide decedents	45% of suicide dececents ages 21–30 Bimodal distribution in the non-suicide decedents: 23% ages 20–25, 23% ages 30–35 Range = 11–45	F	Not Reported	Suicide Death	Autopsy Menstrual status determined by visual examination of the uterine cavity	54.46% of non-pregnant women who died by suicide were menstruating versus 6.75% in the non-suicide decedent group	
Fouriestié et al. 1986[a]	108 suicide attempters	Mean = 25.3 SD = 4.0 Range = Not Reported	F	Assorted (9 with previous psychiatric admission, 15 treated with neuroleptic and/or antidepressant medication, 9 treated with anxiolytics, 15 with previous suicide attempts)	Recent Suicide Attempts	Blood sample within 12 h of attempt: estradiol and progesterone	Suicide attempts were more likely to happen during phases with low estradiol, during the first week of the menstrual cycle (42%) and after the fourth week (12%). Frequency of suicide attempts did not vary significantly during the menstrual cycle in OC users.	Patients admitted to the ICU were excluded.
Gustavsson et al. 2003	43 suicide attempters	Mean = 38.0 SD = 12.0 Range = Not Reported	M	Assorted (14 SUD, 9 DDNOS, 10 MDD, 4 Dysthymia, 9 Adjustment disorder, 4 Anxiety disorder, 2 Psychosis)	Recent Suicide Attempts	CSF in days (5–57 days, mean = 16) following suicide attempt: testosterone	Suicide attempters with depressive disorders showed higher CSF testosterone than those with other psychiatric diagnoses.	CSF testosterone positively correlated with irritability and negatively correlated with social desirability.

Table 1 (continued)

Publication	Sample size and characteristics	Age (years)	Sex	Psychiatric condition	Suicide-related outcome	Study design and methods	Findings	Additional notes
Papadopoulou et al. 2018[a]	70 suicide attempters	Mean = 35.5 SD = 8.9 Range = 18–52	F	Assorted (28 MDD, 13 BD, 14 Psychosis, 15 PD or adjustment disorder)	Recent Suicide Attempts	Blood sample within 72 h of suicide attempt or within 48 h after transfer to the ICU: progesterone, LH, FSH Menstrual status determined with progesterone levels, LH and FSH were used to rule out menopausal status	Suicide attempts were more frequent in the last 4 days of days of luteal phase and during the 4 days of menses.	No effect of menstrual status on lethality (violent vs non-violent mode of attempt) or psychiatric diagnosis.
Markianos et al. 2009[a]	15 suicide attempters (intentional jumps) 18 accident victims (falling from a height) 40 healthy controls	SA: Mean = 39.9, SD = 14.3, Range = 22–62 Non-SA: Mean = 37.6 SD = 15.2 Range = 20–66 HC: Mean = 31.6 SD = 9.0 Range = 25–59	M	Assorted (10 SCZ, 5 MDD)	History of Suicide Attempts	Blood: testosterone, LH, FSH	Suicide attempters had lower levels of testosterone (trending, p = 0.065) and LH compared to accident victims. Both suicide attempters and accident victims had lower levels of testosterone and LH compared to HC.	
Martin et al. 1997[b]	81 female and 79 male adolescents	Mean = 16.0 SD = 1.0 Range = 15–19	Both	Not Reported	History of Suicide Attempts and Suicidal Ideation	Blood: progesterone	In males, progesterone was higher in those with past suicide attempts and with suicide ideation. In females, progesterone levels negatively correlated with past suicide attempts and disclosed suicide ideation	SUD excluded from analyses.
Roland et al. 1986	39 suicide decedents 48 non-suicide (sudden death) decedents	SA: Mean = 39.1 SD = 18.3 Range = 15–76 HC: Mean = 51.5 SD = 13.8 Range = 12–79	M	Not Reported	Suicide Death	Autopsy Blood: testosterone	Suicide decedents showed higher levels of testosterone compared to non-suicide decedents.	
Sher et al. 2012	67 patients with bipolar disorders and at least one past suicide attempt (51 female, 16 male)	Mean = 34.5 SD = 9.9 Range = 18–75	Both	Bipolar Disorder	History of Suicide Attempts	Blood: testosterone	Testosterone levels positively correlated with the number of past suicide attempts, while controlling for sex.	Testosterone levels were also positively correlated with number of manic episodes, while controlling for sex.
Sher et al. 2014	51 patients with bipolar disorder and at least one past suicide attempt	Mean = 33.2 SD = 9.6	F	Bipolar Disorder	History of and Prospective Suicide Attempts (prospective follow-up for up to 2.5 years)	Blood: testosterone	At baseline, testosterone levels positively correlated with the number of suicide attempts and past major depressive episodes. Higher testosterone levels predicted suicide attempts in the follow-up period.	
Sher et al. 2018[a]	17 combat veterans with post-deployment suicide attempt (0 female, 17 male) and 17 non-suicidal combat veterans (2 female, 15 male)	SA: Mean = 37.5 SD = 11.6 Non-SA: Mean = 35.7 SD = 10.8	Both	PTSD	History of Suicides Attempt and Suicidal Ideation	Blood: DHEA, DHEA-S	Suicide attempters had lower levels of DHEA and DHEA-S compared with nonattempters. Suicidal ideation negatively correlated with DHEA and DHEA-S levels across all participants. Suicidal ideation negatively correlated with DHEA-S levels in nonattempters.	DHEA/DHEA-S ratios positively correlate with adolescent and adult aggresion scores in suicide attempters.

Table 1 (continued)

Publication	Sample size and characteristics	Age (years)	Sex	Psychiatric condition	Suicide-related outcome	Study design and methods	Findings	Additional notes
Stefansson et al. 2016	28 suicide attempters (10 female, 18 male) 19 healthy controls (7 female, 12 male)	SA: Mean = 44.0 SD = 14.6 Range = 23–66 HC: Mean = 30.0 SD = Not Reported Range = 23–48	Both	Assorted (MDD, PTSD, SUD)	Recent Suicide Attempts and Prospective Suicide Death (prospective 21-year follow-up)	CSF and blood in days (mean = 8.6, range = 2–17 days) following suicide attempt: testosterone, cortisol	In males, CSF and blood testosterone levels were higher in suicide attempters compared to healthy controls. In females, no differences.	In males, CSF testosterone/cortisol ratio positively correlated with impulsivity and aggressiveness in the suicide attempters. No differences associated with MDD, PD, or SUD
Tripodianakis et al. 2006	80 suicide attempters (29 with schizophrenia) 56 healthy controls 29 nonattempters with schizophrenia	SA: Mean = 34.4 SD = 12.6 HC: Mean = 35.3 SD = 8.7	M	Schizophrenia	History of Suicide Attempts	Blood: testosterone, LH, FSH	Suicide attempters had lower blood testosterone compared to healthy controls. Attempters with schizophrenia had lower levels of testosterone compared to nonattempters with schizophrenia.	Attempters who used a violent method had lower testosterone levels than non-violent attempters.
Zhang et al. 2015	245 suicide attempters (172 female, 73 male) 245 healthy controls (172 female, 73 male)	SA: Mean = 42.9 SD = Not Reported Range = 16–50 HC: Mean = 37 SD = Not Reported Range = 14–53	Both	Not Reported	History of Suicide Attempts	Blood: testosterone	In males, testosterone was higher in male suicide attempters compared to healthy controls. In females, no significant differences.	

F female, *FH* follicular hormone, *DDNOS* depressive disorder not otherwise specified, *ICU* intensive care unit, *LH* lutenizing hormone, *M* male, *MDD* major depressive disorder, *OC* oral contraceptive, *OD* overdose, *OCD* obsessive-compulsive disorder, *PTSD* post-traumatic stress disorder, *SA* suicide attempt, *SCZ* Schizophrenia, *SD* standard deviation, *SUD* substance use disorder.

[a]Low-quality study due to limited sample size and/or limitations in study design (e.g., single-timepoint cross-sectional associations between ovarian hormones and suicidal thoughts and behaviors).

[b]Adolescent sample.

neurobiological sensitivity to *changes* in estradiol and progesterone—rather than between-person differences in levels of these hormones—that represents a key trait-like source of variance for understanding suicide risk in the context of puberty, a time when the stability of these hormone dynamics is still in flux [47]. Intensive (e.g., daily samples) longitudinal studies conducted in adult women have also provided little evidence that hormone levels at a single point in time are correlated with levels within the same cycle and also for subsequent cycles, particularly for estradiol [48, 49]. It is also important to note that both between-person and within-person variability of ovarian hormone levels are affected by multiple confounds that investigators ought to consider, including, but not limited, to cycle length [50], diurnal effects [51], cycle phase [48, 49], effects of study participation [48], anovulation [52], culture and/or diet [53], and personal and family medical history (e.g., polycystic ovary syndrome, breast cancer [48, 54]). Nevertheless, dense-sampling studies have demonstrated robust between-person effects of menstrual phase, such that progesterone is reliably higher in the luteal phase relative to the follicular phase and, albeit to a lesser extent, estradiol is higher in the follicular phase relative to the luteal phase [48, 49]. We note in Table 1 which studies employ non-experimental cross-sectional designs and that attempt to relate ovarian hormones without consideration of these relevant factors (e.g., cycle phase).

Given these limitations in measuring ovarian hormones, it is not surprising that little is known about the effects of estradiol and progesterone on the human brain in general, much less on the circuits we have hypothesized are implicated in STB risk. Nonetheless, there is evidence that estrogen receptors are expressed strongly in brain regions involved in social cognition broadly and social rejection specifically, including the amygdala, hippocampus, ACC, and vmPFC [55–58]. Several studies have also found that women with comparably higher levels of estradiol (endogenous or synthetic) show greater amygdala-based resting-state functional connectivity and activation in ACC and vmPFC, which are key regions involved in processing salient information [59–61]. Similarly, longitudinal studies with naturally cycling women have documented larger hippocampal gray matter volumes during periods of relatively high estradiol (late follicular/preovulatory phases) than of relatively low estradiol early follicular/premenstrual phases [62, 63].

There is also evidence from functional magnetic resonance imaging (fMRI) studies wherein groups of women during different phases of their menstrual cycle are compared that show that relatively higher levels of estradiol (pre-ovulatory phase) are associated with greater activation of the hippocampus during both cognitive tasks [64] and affective stress tests [65], and with stronger functional connectivity of the hippocampus with brain regions involved in processing salient information [63]. Similarly, higher levels of progesterone (luteal phase) have also been found to be associated with activation of the striatum and PFC during cognitive processing [64]. Finally, results from fMRI studies utilizing dense-sampling designs of naturally cycling women have reported intriguing, albeit inconsistent, results. One study found no effects of estradiol across the menstrual cycle on intrinsic connectivity patterns [66], whereas another found that higher levels of estradiol drove stronger subsequent connectivity within attention networks (specifically among brain circuits that are implicated in internally focused attentional states and externally focused attentional states) [67]. In contrast, one study found that progesterone mediated patterns of positive functional connectivity between the hippocampus and PFC [66] whereas another study found that higher levels of progesterone was associated with lower connectivity across all networks [67]. It is clear that additional research is needed to clarify the magnitude and directionality of the effects of estradiol and progesterone on patterns of brain connectivity, particularly in adolescents, and their effects on longer-term neurodevelopmental trajectories. It is also critical that researchers work to characterize the extent to which these hormones effect brain circuits specific to STBs, or alternatively, whether they are implicated in mental illness risk more broadly. Nevertheless, the studies we have reviewed provide initial evidence that brain circuits that support the regulation of affectively salient stimuli are sensitive to the effects of estradiol and progesterone; importantly, these same circuits have also been shown to be affected in adults and adolescents with STBs [28, 68].

Additional research is also needed to elucidate the precise neural mechanisms by which estradiol may impact the development of brain circuits, including its effects on neurotransmitters. Estradiol has also been found to alter serotonin transmission, binding, and metabolism by increasing the production of tryptophan and inhibiting the expression of the serotonin reuptake transporter gene (for a review, see [69]). Serotoninergic abnormalities in the number of serotonergic neurons and in serotonin transportation, receptor binding, and levels have all been found in victims of suicide [70, 71] (for a review, see [72]). Serotonin has been shown to be associated with cortisol reactivity to stressors [73] and positively associated with greater 5-HT_{1A} receptor binding—which could contribute to lower serotonin signaling by inhibiting further serotonin release into the synapse—in depressed patients who died by suicide versus both depressed and psychiatrically healthy individuals who did not die by suicide [74]. Similarly, allopregnanolone, a progesterone metabolite, binds to $GABA_A$ receptors, which mediates the majority of inhibitory signaling in the brain [75] and plays an important role in downregulating the

HPA-axis in response to acute stressors [76]. Several studies have found that stress exposure alters the availability of $GABA_A$ receptors as well as their composition and sensitivity to neurosteroid regulation, which, in turn, can influence subsequent responses to stress [77–79]. Emerging data has also implicated GABA dysfunction in STBs, whereby depressed patients who died by suicide showed a higher expression of genes that encode for proteins involved in GABAergic synaptic transmission in the ACC and a lower expression of these genes in the dorsolateral PFC than did both depressed and psychiatrically healthy individuals who did not die by suicide [80]. Future work is needed to explicitly investigate the extent to which patterns of structural or functional connectivity of circuits involving the amygdala, hippocampus, ACC, and PFC that are developing in response to puberty-related changes in estradiol and progesterone (and related neurosteroids) exhibit corresponding changes in neurotransmitter systems that support socioemotional processes relevant to STB risk.

Testosterone

Unlike ovarian hormones, which have been studied primarily in female adolescents, investigators have found in both sexes that levels of androgens are associated with appetitive behaviors, aggression, competition, and other related social behaviors that are relevant to STB risk [81, 82]. In female adolescents, the sex steroid with the most androgenic activity is DHEA, which is produced by the adrenal cortex (discussed separately in the following section); in males adolescents, it is testosterone [31, 32].

In contrast to the data published thus far on ovarian hormones, there is evidence of strong within-person reliability and stability in testosterone levels in both men and women. Indeed, some researchers contend that levels of testosterone are trait-like [83]. Perhaps not surprising, testosterone has been linked more strongly with STB risk, and specifically with suicide attempts, than has any other pubertal hormone (see Table 1). However, the empirical findings thus far have been mixed (for opposing reviews, see [84, 85]). Several researchers have theorized that testosterone is linked to suicidal behaviors through its modulation of emotion-related impulsivity and impulsive aggression, which are considered to be among the most robust predisposing traits to suicidal behaviors in youth [86]. In drawing from the literature of studies in which testosterone is administered and behaviors are subsequently assessed, as well as studies that link changes in social interactions with changes in endogenous testosterone, there appear to be reliable effects of testosterone on socially motivated behaviors, including exerting dominance and displays of aggression (either physically or non-physically) and other social status seeking behaviors (for a review, see [87]). While testosterone exert complex effects on interpersonal behavior, longitudinal studies show that the puberty-related increases in testosterone are not accompanied by a concurrent rise in aggressive behaviors [83, 88, 89]. Additional studies in animals as well as in humans [90] suggest that testosterone levels correlate more closely with social dominance, rather than aggressive behaviors. Hence, testosterone may be an important moderator of the behavioral response to events associated with loss of social status [91], which are known precipitants of STBs. In the context of adolescent STB risk, it may very well be that testosterone is a driver of heightened sensitivity to social context, which can lead to significant emotion dysregulation and impulse control and, in turn, elevated STB risk.

Consistent with these points is evidence that elevated testosterone has also been found in male adults to be associated with psychological states and individual traits associated with suicide risk, including depression severity, irritability, and impulsive aggression [92, 93]. Researchers have found higher levels of testosterone in both male and female suicide attempters than in their same-sex non-attempting counterparts [94], and in post-mortem samples of suicide completers compared to individuals who died from other causes [95]. In a recent study, male, but not female, young adults who attempted suicide had higher levels of testosterone than did age- and sex-matched healthy volunteers [92]. Thus, there studies do not preclude the possibility that testosterone levels are higher in individuals with a mental disorder than they are in healthy persons. One study found in women with bipolar disorder with a history of suicide attempt that higher testosterone levels predicted subsequent suicide attempts [96]. Interestingly, in another sample of men and women with bipolar disorder, higher levels of testosterone were associated with number of suicide attempts only after controlling for sex [97]. In contrast, other studies have found lower levels of testosterone both in male suicide attempters relative to psychiatric healthy controls [98] and in patients who were hospitalized due to accidents [99]. One possible discrepancy for these findings involves the measurement of testosterone in plasma versus cerebrospinal fluid; studies assaying plasma testosterone have typically reported inverse associations with suicide attempts. Nevertheless, it is clear that adequate clinical controls (attempters or completers relative to nonattempters who otherwise have similar clinical histories) must be included in research in order to clarify the results of these studies.

As is the case with research on ovarian hormones and STBs, no studies have examined the psychobiological mechanisms and specific brain circuits through which testosterone may drive risk for STBs in either adolescents or adults. Testosterone influences many of the same neurotransmitters as does estradiol [25]; moreover, testosterone

can be converted into estradiol via aromization [100], adding additional complexity in these processes should be addressed in future work. That said, one pathway through which testosterone may affect mood and cognition is through altering dopaminergic neurotransmission: several studies have found evidence of differences between suicide attempters and completers in dopamine receptor density, transporter binding capacity, and metabolism [101–103], primarily in the striatum. In fact, testosterone receptors are widely expressed in dopaminergic neurons of the striatum and portions of the PFC, including vmPFC and OFC [104]. Research with both macaques and humans has documented changes in the mechanisms that underlie dopamine signaling during adolescence, which may be explained, in part, by effects of testosterone (and estrogen; for reviews, see [105, 106]). In particular, testosterone has been shown to increase basal dopamine levels and decrease the number of enzymes that break down dopamine in the striatum and PFC [105]. Consequently, it is possible that adolescent-typical increases in testosterone during puberty, especially in boys, contribute to elevated STB risk through enhanced dopaminergic transmission. This is consistent with previous reports that, compared to depressed patients who did not die by suicide, depressed patients who die by suicide have lower growth hormone responses to apomorphine [102, 107], which is an indication of higher dopamine transporter binding [108]. Nonetheless, it is important to note that some studies provide evidence of lower dopamine (measured by lower levels of homovanillic acid and total dopamine in urine) in patients who attempted suicide versus those who did not, see [103], or of no significant associations between dopamine levels/receptor binding and STBs [109].

With respect to brain development, higher levels of pubertal testosterone have been associated with white matter organization in tracts implicated in social cognition and emotion regulation, including the uncinate fasciculus (which connects the amygdala and vmPFC) and corpus callosum in both boys [110, 111] and girls [29, 112]. Several researchers have documented other effects of testosterone on adolescent and young adult brain and behavior. Human neuroimaging studies using fMRI have shown that in healthy young men, activation in the amygdala—a brain structure that is rich in androgen receptors [63] and is affected by circulating androgens [64]—to fearful and angry faces co-varies positively with individual differences in serum testosterone concentrations [65, 66] (but see [67] for opposing results). In an experimental study, administration of testosterone was associated with increased amygdala reactivity to threat-related stimuli in young women [113]. These results are consistent with a longitudinal study in adolescents that found that increased levels of pubertal testosterone disrupted typical coupling between the amygdala and OFC, leading to increased amygdala reactivity to threat-related stimuli in both sexes [114]. Finally, whereas in boys higher testosterone levels are associated with lower activation in the striatum and PFC during processing of emotional conflict, in girls higher testosterone levels are associated with lower activation in the PCC/precuneus [115]. Other investigators have shown in both boys and girls that higher levels of testosterone are associated with higher striatal activation during reward consumption [116], and with higher striatal activation when adolescents select smaller yet more immediate rewards [117]. In an observational longitudinal neuroimaging study in male and female adolescents between the ages of 8–27 years, activation in the striatum (nucleus accumbens) to rewarding stimuli peaked during adolescence, and was associated with accompanying changes in testosterone levels [118]. In a recent study, in which salivary testosterone was measured in adolescents before and after a neuroimaging scan, acute increases in testosterone were associated with smaller differences in activation between reward cues signaling reward or non-reward outcomes for a given trial in vmPFC and posterior cingulate cortex (PCC) in both sexes [119]. Thus, by altering neural processing of both negative (e.g., threatening) and positive (e.g., rewarding) valenced stimuli, testosterone may be facilitating the expression of adolescent sensation-seeking and risk-taking behavior [118] through dopaminergic transmission among these subcortical and prefrontal circuits. Certainly, these processes may not be specific to STBs and may represent a more general diathesis for mental disorders that are characterized by sensation-seeking and risky behaviors; however, future research is needed to test this possibility more explicitly.

In sum, studies to date have demonstrated that higher levels of testosterone are associated with aspects of emotion-related impulsivity in males with a history of suicidal behaviors (e.g., attempts). Higher levels of testosterone are also associated with steeper temporal discounting of rewards driven by patterns of activation in striatal, vmPFC, and OFC in both sexes; sex differences in the effect of testosterone on adolescent brain circuits appear to be most prominent in emotion-related contexts. Indeed, in adults, both endogeneous levels [120] and exogenous administration [121] of testosterone enhance risk taking. The predisposition to take risky decision has been shown to be a critical risk factor for suicidal behavior in adults [122] and, to a lesser extent given the sparse literature on this topic, in adolescents [123]. Hence, functional alterations in decision-making systems induced by puberty-related rises in testosterone levels, or their context-dependent fluctuations, might be related to suicidal risk through an increase probability of perpetrating highly risky and self-destructive behaviors in the face of overwhelming stress (particularly social stressors involving loss of desired status). Future work is needed to test explicitly whether higher levels of testosterone exacerbate affect

striatal and PFC systems underlying maladaptive responses to social stressors that, in turn, lead to heightened risk for STBs.

DHEA

DHEA (and its sulfate, DHEA-S) is the most abundant steroid hormone in humans and is a precursor to sex steroids. Researchers have not yet examined the relation between DHEA (or DHEA-S) and STBs in adolescents; moreover, and the few studies that have been conducted with adults differ in their measurement of DHEA. One study showed that male suicide attempters had higher levels of DHEA-S than did healthy controls, and that exposure to extreme social threat (i.e., interpersonal violence) as a child was negatively correlated with the ratio of cortisol/DHEA-S [124]. Among adults diagnosed with PTSD, those who had attempted suicide had significantly higher levels of DHEA than did those without a history of attempt [125]. Another study found that, in combat veterans, the ratio of DHEA/DHEA-S was positively correlated with total adolescent aggression scores, total adulthood aggression scores, and lifetime aggression scores in those who had attempted suicide but not in nonattempters [126].

Overall, these studies suggest that higher levels of DHEA or DHEA-S may be associated with risk for STBs. Indeed, DHEA modulates neurotransmitter systems implicated in suicidal thoughts and mood disturbances, including acting as an antagonist for GABA receptors, with genes widely expressed in hippocampus, amygdala, and striatum [127]. There is also intriguing evidence that psychiatric conditions characterized by excessive stress and elevated suicide risk lead to a downregulation of neurosteroid biosynthesis—including the conversion of DHEA to GABAergic metabolites, such as allopregnanolone [128]—and changes in $GABA_A$ receptor subunit composition [129]. Thus, STB risk may be associated with higher levels of DHEA due to an insufficient ability to metabolize DHEA. With respect to the effects of DHEA on the adolescent brain, however, there have been a small number of studies. Thus, far investigators have found that higher levels of DHEA are associated with larger hippocampal volumes in both male and female adolescents [130] and with lower white matter organization across a broad set of white matter tracts [110]. Other studies have found that higher DHEA levels are associated with reduced cingulate activation and greater functional connectivity between the amygdala with ACC and other regions involved in processing salient information in adolescents during the processing of social stimuli (e.g., viewing fearful faces) [131]. Moreover, higher DHEA levels are associated in girls with lower activation in cingulate regions implicated in processing salience information and with greater externalizing problems [132], but in boys with stronger functional connectivity between the amygdala and inferior frontal gyrus (the opposite pattern was found in girls) and with higher anxiety symptoms [131].

Despite the heterogeneity in clinical, developmental, and demographic features across the studies reviewed in this section, there appear to be consistent associations between higher levels of DHEA and both STBs and altered structure and function of brain circuits underlying emotion regulation. In adults with clinical disorders, higher levels of DHEA (either alone or relative to cortisol) appear to be associated with STBs. In psychiatrically healthy adolescents, higher levels of DHEA are associated with reduced downregulation of affective signals in the amygdala from emotion regulatory cortical regions, including the ACC and portions of the PFC. Thus, higher levels of DHEA may disrupt the development of emotion regulatory brain circuits across adolescence in ways that increase risk for STBs when individuals are exposed to interpersonal stressors (particularly those characterized by threat). However, it is worth noting one study of psychiatrically healthy adults in which administration of exogenous DHEA reduced activation in the amygdala and hippocampus, increased activation in the vmPFC, and led to stronger connectivity between the amygdala and hippocampus during an emotion reappraisal task; moreover, decreased activation in the hippocampus during the task was associated with lower negative affect, suggesting that higher levels of DHEA improve negative mood by downregulating affective signals in the hippocampus [133]. Additional research is needed to examine whether these associations hold in clinical samples of adolescents and whether (or how) they are related to risk for STBs.

Consideration of sex differences

Across most countries, being female increases the risk of suicidal thoughts and related behaviors [5]. Despite this higher prevalence of STBs in women, men are more likely to die by suicide [134], leading to "the gender paradox of suicide." It is notable that the sex difference in suicide deaths increases dramatically in adolescence [135], suggesting that puberty plays an important role in explaining this difference (although this sex difference is now narrowing among adolescents [136], indicating that other factors, including environmental or cultural influences, are also likely). Whereas past-year ideation, plans, and attempts tend to peak during mid-adolescence (~16 years) in girls, these rates increase steadily throughout mid- to late adolescence in boys [137, 138].

As in adults [5, 139], there are sex differences in suicide-related outcomes in adolescents that are mediated in part by differences in lethality and mental illness. In a psychological autopsy study, adolescent male suicide victims were

more likely to use lethal methods and had a higher prevalence of conduct disorder than did female victims [140]. A recent meta-analysis of sex-specific suicide risk in adolescents (ages 12–26 years) found distinct clinical and environmental risk factors for suicide attempt in male and female adolescents [139]. Thus, in addition to the hormonal differences that underlie sex-specific neurodevelopmental changes, different clinical conditions (externalizing disorders in boys versus mood disorders in girls) and social stressors (peer influence in boys versus direct trauma/victimization from romantic relationships in girls) may further explain or moderate these pathways to risk for STBs. Because of the sparse literature, however, it is not clear whether sex-specific risk factors are present before puberty, whether puberty affects neuroendocrine systems in sex-specific ways to increase risk for STBs, or whether boys and girls are exposed to differential environmental risk factors as a result of going through puberty. Future research should attempt to clarify the extent to which pubertal processes play a central role in STBs or whether they are implicated in mental illness more generally.

While adrenal and gonadal hormones and physical maturation are important indicators of pubertal development, other aspects of puberty may be relevant in the context of adolescent risk for STBs. Pubertal timing—the age at which individuals mature relative to their peers [141]—has been linked to individual differences in mental disorders in adolescence [142, 143]. Moreover, the timing of pubertal onset may have a different impact on developing neuroendocrine function depending on whether it occurs earlier or later in a given individual [34, 139]. Given that girls typically enter puberty earlier than do boys, considering the role of pubertal timing may also elucidate sex differences in risk for STB—as well as mental illness more generally—and accompanying endocrine and neural processes. A growing number of studies are reporting that early menarche is associated with elevated risk for suicidal ideation in adolescent girls [144–146]. In a recent longitudinal study of a large birth cohort, earlier age of menarche was associated with increased suicidal behaviors at 16 and 21 years of age [147], suggesting that earlier puberty has an enduring effect on STB risk throughout adolescence and young adulthood. In addition, considerable evidence suggests that pubertal timing is influenced by early life adversity [148], which itself is a robust predictor of STBs [149]. Thus, early puberty resulting from early adversity may be a mechanism by which suicide risk is instantiated or exacerbated in vulnerable adolescents. Finally, it is important to acknowledge that sexual orientation and identification as a sexual minority are increasingly being recognized as risk factors for STBs [150]; more research in this area is critically needed.

Future directions

No research has yet examined whether and how pubertal hormones affect neurodevelopmental trajectories of brain circuits that mediate social cognition and emotion-related impulsivity explain risk for STBs during adolescence. Studies are needed to clarify the role of pubertal timing and the multidimensional mechanisms—biological, social, cultural—by which puberty-related processes influence risk for STBs. An important next step for the field is to first establish reliable associations between pubertal hormones and adolescent brain structure and function and to then map those associations onto neural circuits underlying STBs in adolescents. Other critical knowledge gaps include disentangling neuroendocrine mechanisms that are more closely linked to suicidal thoughts versus attempts (and other self-harming behaviors) and that facilitate the transition from ideation to action [151].

Although we focused here on the effects of pubertal hormones, we should acknowledge that there is considerable evidence that HPA-axis dysfunction is associated with STBs and self-harming behaviors in adolescents [13, 21, 152, 153]. Puberty is also the time when HPG-axis regulation of the HPA axis develops [154–157]. Given that the expected suppression of the HPA axis by the HPG axis [158] is disrupted in individuals at risk for STBs (due to such factors altering developmental changes in stress response and regulatory systems, including experiences of adversity and exposure to trauma during early life [154]), it is critical that researchers explicitly examine the role of puberty in this context.

It is also important that researchers in this area consider moving away from a nomothetic framework for the prediction of STBs and instead adopting idiographic or person-centered models. Puberty is a highly individualized process; adolescents differ markedly in their hormone levels, their neurobiological sensitivity to typical endocrine changes [14], their pubertal timing, their neurodevelopmental trajectories, and their psychological responses to maturation. Certainly, measuring these variables at the individual level is challenging and will have to take into account population norms on several dimensions (e.g., ethnicity, socio-demographic factors). Absent such data, standardization within age bands or residualized scores (e.g., regressing pubertal stage on age to obtain a measurement of pubertal timing) can be used to capture individual variability (for a treatment of these issues, see [159]). Another solution is to use "dense sampling" approaches that leverage repeated assessments (e.g., self-reported mood states, saliva samples, neurobiological measurements) from the same individual in order to capture intra-individual variation in these processes [13], which we argue is critical for understanding the contribution of ovarian hormones in risk for STBs in adolescent

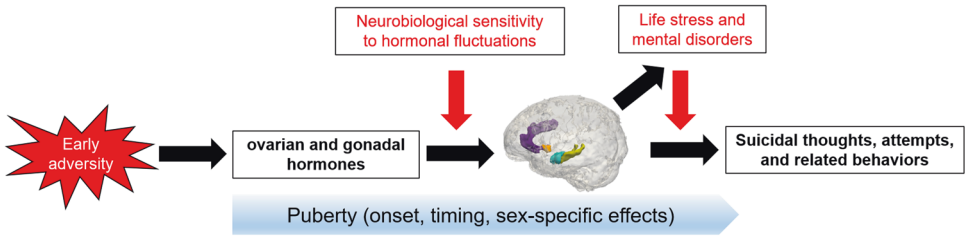


Fig. 2 Conceptual model linking aspects of the pubertal transition with risk for suicidal thoughts and behaviors. Experiences of early adversity affect the programming and development of endocrine and neural systems which undergo significant maturation during puberty. Puberty-related changes in ovarian, gonadal, and other related hormones shape the neural circuits underlying social cognition, emotion regulation, and impulse control (which include structures such as the amygdala, hippocampus, striatum, anterior cingulate cortex, and portions of prefrontal cortex). Alterations in these circuits may partially explain the ways in which changes in sex hormones are linked with the emergence of suicidal thoughts and behaviors during adolescence. Moderators of these processes, including a neurobiological sensitivity to ovarian hormones, experience of ongoing life stressors, and underlying mental disorders, are highlighted in red.

females. Dense-sampling has the additional advantage of being able to identify proximal factors that lead to suicidal states and other related high-risk events, which are often transient and rarely captured in the laboratory or clinic [160].

Other important study design considerations include larger sample sizes of both sexes in order to detect sex differences more reliably. As we alluded to in the earlier sections of our review, there are other important confounds to consider that will affect the reliability and validity of estimating the effects of within-person changes in sex hormones on brain and behavior. These confounds include genetic factors, environmental influences, medical conditions, and cycle length (given that the range of menstrual cycles and periods of anovulation are longer in adolescents than in adults, particularly in the first few years post-menarche [161]). Finally, understanding the normative associations between changes in hormones with the structural and functional development of the brain is a critical next step that must be established before understanding how these alterations are altered in individuals at risk for STBs. Gaining a more comprehensive understanding of neurobiological trajectories and the mechanisms of neuroplasticity in adolescents as a consequence of puberty will also contribute to our knowledge of sex-specific psychiatric interventions and may help to address the heterogeneity of risk factors identified to date [5]. For instance, the granularity of data obtained from dense-sampling methods will facilitate the identification of individuals with hormonal sensitivity to the typical fluctuations occurring during puberty, as well as characterize within-person pubertal processes associated with responses to social stressors and poorer impulse control during heightened emotional conflict. With this information, researchers and clinicians will be able to stratify individuals on the basis of neuroendocrine risk (e.g., those with hormone sensitivity) and also to identify points in time (e.g., dramatic increases in stress perception or depressive symptoms, unusual variability in DHEA) during which an individual might benefit from immediate intervention.

Summary and conclusions

Globally, suicide has surpassed all physical diseases as a cause of death in adolescents. Puberty drives psychobiological changes in adolescence that have not been examined explicitly in relation to suicide risk. In this review, we conceptualized suicidal thoughts and behaviors in adolescents as resulting from pathological responses to social stressors at a time when stress-regulatory systems are still maturing. We argued that alterations in brain circuits—comprised of connections among the amygdala, hippocampus, striatum, ACC, vmPFC, and OFC—that underlie social cognition, emotion regulation, and impulse control and are shaped by puberty-related changes in sex hormones. We propose that alterations in these circuits may partially explain the ways in which changes in sex hormones are linked with increased suicidal thoughts and behaviors during adolescence. However, we also highlighted critical moderators to be considered in this model, including a neurobiological sensitivity to fluctuations in ovarian hormones, exposure to early adversity, and underlying mental disorders (Fig. 2). To date, these specific hypotheses have not been tested. Although there is emerging research identifying shared and unique neural circuits that underlie suicidal thoughts and behaviors [68], it is paramount that that these associations be examined in prospective studies. It will also be important to use both experimental designs and large-scale longitudinal studies to elucidate the extent to which pubertal hormones affect the acute functioning of these circuits and drive the development of these circuits over the course of adolescence and young adulthood. Moreover, we expect that although pubertal hormones are not a primary driver of suicide risk, they may play an outsized role in individuals with a neurobiological sensitivity to hormonal fluctuations. Therefore, it is critical that we identify key moderators of the paths in our model, which we hypothesize includes neurobiological sensitivity to hormonal fluctuations, experiences of adversity and life

stress that influence neurodevelopmental trajectories (which, in turn, may result in generating and experiencing additional stress), and underlying mental disorders. In addition to conducting longitudinal studies with larger sample sizes, we suggest that researchers also use dense-sampling methods to identify individuals according to these stratification parameters (e.g., hormonal sensitivity), as well as points in time at which individuals may be at risk and could benefit from more immediate intervention. In conclusion, we want to emphasize that increasing our understanding of pubertal science across endocrine, neural, and psychosocial domains will yield significant insights concerning how best to reduce the frequency of suicide-related deaths during the vulnerable period of adolescence.

Acknowledgements This work was supported by the National Institute of Mental Health (R37MH101495 to IHG, K01MH117442 to TCH), the Fonds de Recherche du Québec—Santé (FRQS/MSSS Resident Physician Health Research Career Training Program to AJG), Precision Health and Integrated Diagnostics Center at Stanford (PHIND to IHG and TCH), and the Ray and Dagmar Dolby Family Fund (to TCH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. All authors report no biomedical conflicts of interest. The funding agencies played no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Mokdad AH, Forouzanfar MH, Daoud F, Mokdad AA, El Bcheraoui C, Moradi-Lakeh M, et al. Global burden of diseases, injuries, and risk factors for young people's health during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2016;387:2383–401.
2. Murphy SL, Mathews TJ, Martin JA, Minkovitz CS, Strobino DM Annual Summary of Vital Statistics: 2013–2014. Pediatrics [Internet]. 2017 [cited 2020 Apr 10];139. Available from: https://pediatrics.aappublications.org/content/139/6/e20163239
3. Curtin SC, Heron M. Death rates due to suicide and homicide among persons aged 10–24: United States, 2000–2017. NCHS Data Brief. 2019;352:1–8.
4. Carter G, Milner A, McGill K, Pirkis J, Kapur N, Spittal MJ. Predicting suicidal behaviours using clinical instruments: Systematic review and meta-analysis of positive predictive values for risk scales. Br J Psychiatry. 2017;210:387–95.
5. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull. 2017;143:187–232.
6. Cha CB, Franz PJ, Guzmán EM, Glenn CR, Kleiman EM, Nock MK. Annual research review: suicide among youth – epidemiology, (potential) etiology, and treatment. J Child Psychol Psychiatry. 2018;59:460–82.
7. Glenn CR, Kleiman EM, Kellerman J, Pollak O, Cha CB, Esposito EC, et al. Annual research review: a meta-analytic review of worldwide suicide rates in adolescents. J Child Psychol Psychiatry. 2020;61:294–308.
8. Nelson EE, Leibenluft E, McClure EB, Pine DS. The social re-orientation of adolescence: a neuroscience perspective on the process and its relation to psychopathology. Psychol Med. 2005;35:163–74.
9. Paus T, Keshavan M, Giedd JN. Why do many psychiatric disorders emerge during adolescence? Nat Rev Neurosci. 2008;9:947–57.
10. Rudolph KD Puberty as a Developmental Context of Risk for Psychopathology. In: Lewis M, Rudolph KD, editors. Handbook of Developmental Psychopathology. Boston, MA: Springer US; 2014. p. 331–54.
11. O'Connor RC, Nock MK. The psychology of suicidal behaviour. Lancet Psychiatry. 2014;1:73–85.
12. Lutz P-E, Mechawar N, Turecki G. Neuropathology of suicide: recent findings and future directions. Mol Psychiatry. 2017;22:1395–412.
13. Miller AB, Prinstein MJ. Adolescent suicide as a failure of acute stress-response systems. Annu Rev Clin Psychol. 2019;07:425–50.
14. Owens SA, Eisenlohr-Moul T. Suicide risk and the menstrual cycle: a review of candidate rdoc mechanisms. Curr Psychiatry Rep. 2018;20:106.
15. King CA, Merchant CR. Social and interpersonal factors relating to adolescent suicidality: a review of the literature. Arch Suicide Res. 2008;12:181–96.
16. Liu RT, Miller I. Life events and suicidal ideation and behavior: a systematic review. Clin Psychol Rev. 2014;34:181–92.
17. Stewart JG, Shields GS, Esposito EC, Cosby EA, Allen NB, Slavich GM, et al. Life stress and suicide in adolescents. J Abnorm Child Psychol. 2019;47:1707–22.
18. Platt B, Cohen Kadosh K, Lau JYF. The role of peer rejection in adolescent depression. Depress Anxiety. 2013;30:809–21.
19. Prinstein MJ, Giletta M Peer relations and developmental psychopathology. In: Developmental psychopathology: theory and method, Vol 1, 3rd ed. Hoboken, NJ, US: John Wiley & Sons Inc; 2016. p. 527–79.
20. Whiteside SP, Lynam DR, Miller JD, Reynolds SK. Validation of the UPPS impulsive behaviour scale: a four-factor model of impulsivity. Eur J Personal. 2005;19:559–74.
21. Eisenlohr-Moul TA, Miller AB, Giletta M, Hastings PD, Rudolph KD, Nock MK, et al. HPA axis response and psychosocial stress as interactive predictors of suicidal ideation and behavior in adolescent females: a multilevel diathesis-stress framework. Neuropsychopharmacology 2018;43:2564–71.

22. van Geel M, Vedder P, Tanilon J. Relationship between peer victimization, cyberbullying, and suicide in children and adolescents: a meta-analysis. JAMA Pediatr. 2014;168:435–42.
23. Auerbach RP, Stewart JG, Johnson SL. Impulsivity and suicidality in adolescent inpatients. J Abnorm Child Psychol. 2017;45:91–103.
24. Mahfouz A, Lelieveldt BPF, Grefhorst A, van Weert LTCM, Mol IM, Sips HCM. et al. Genome-wide coexpression of steroid receptors in the mouse brain: identifying signaling pathways and functionally coordinated regions. Proc Natl Acad Sci USA. 2016;113:2738–43.
25. McEwen BS, Milner TA. Understanding the broad influence of sex hormones and sex differences in the brain. J Neurosci Res. 2017;95:24–39.
26. Kaczkurkin AN, Raznahan A, Satterthwaite TD. Sex differences in the developing brain: insights from multimodal neuroimaging. Neuropsychopharmacology 2019;44:71–85.
27. Scheinost D, Finn ES, Tokoglu F, Shen X, Papademetris X, Hampson M, et al. Sex differences in normal age trajectories of functional brain networks. Hum Brain Mapp. 2015;36:1524–35.
28. Schmaal L, van Harmelen A-L, Chatzi V, Lippard ETC, Toenders YJ, Averill LA, et al. Imaging suicidal thoughts and behaviors: a comprehensive review of 2 decades of neuroimaging studies. Mol Psychiatry. 2020;25:408–27.
29. Ho TC, Colich NL, Sisk LM, Oskirko K, Jo B, Gotlib IH. Sex differences in the effects of gonadal hormones on white matter microstructure development in adolescence. Dev Cogn Neurosci. 2020;42:100773.
30. Vijayakumar N, Op de Macks Z, Shirtcliff EA, Pfeifer JH. Puberty and the human brain: Insights into adolescent development. Neurosci Biobehav Rev. 2018;92:417–36.
31. Reiter EO, Fuldauer VG, Root AW. Secretion of the adrenal androgen, dehydroepiandrosterone sulfate, during normal infancy, childhood, and adolescence, in sick infants, and in children with endocrinologic abnormalities. J Pediatr. 1977;90:766–70.
32. Søeborg T, Frederiksen H, Mouritsen A, Johannsen TH, Main KM, Jørgensen N. et al. Sex, age, pubertal development and use of oral contraceptives in relation to serum concentrations of DHEA, DHEAS, 17α-hydroxyprogesterone, Δ4-androstenedione, testosterone and their ratios in children, adolescents and young adults. Clin Chim Acta. 2014;437:6–13.
33. Dorn LD, Dahl RE, Woodward HR, Biro F. Defining the boundaries of early adolescence: A user's guide to assessing pubertal status and pubertal timing in research with adolescents. Appl Dev Sci. 2006;10:30–56.
34. Schulz KM, Sisk CL. The organizing actions of adolescent gonadal steroid hormones on brain and behavioral development. Neurosci Biobehav Rev. 2016;70:148–58.
35. Piekarski DJ, Johnson CM, Boivin JR, Thomas AW, Lin WC, Delevich K, et al. Does puberty mark a transition in sensitive periods for plasticity in the associative neocortex? Brain Res. 2017;1654:123–44.
36. Juraska JM, Willing J. Pubertal onset as a critical transition for neural development and cognition. Brain Res. 2017;1654:87–94.
37. Saunders KEA, Hawton K. Suicidal behaviour and the menstrual cycle. Psychol Med. 2006;36:901–12.
38. Osborn E, Brooks J, O'Brien PMS, Wittkowski A Suicidality in women with premenstrual dysphoric disorder: a systematic literature review. Arch Womens Ment Health [Internet]. 2020 Sep 16 [cited 2021 Mar 13]; Available from: https://doi.org/10.1007/s00737-020-01054-8
39. Shams-Alizadeh N, Maroufi A, Rashidi M, Roshani D, Farhadifar F, Khazaie H. Premenstrual dysphoric disorder and suicide attempts as a correlation among women in reproductive age. Asian J Psychiatry. 2018;31:63–6.
40. Wei S-M, Schiller CE, Schmidt PJ, Rubinow DR. The role of ovarian steroids in affective disorders. Curr Opin Behav Sci.2018;23:103–12.
41. Schmidt PJ, Martinez PE, Nieman LK, Koziol DE, Thompson KD, Schenkel L, et al. Exposure to a change in ovarian steroid levels but not continuous stable levels triggers PMDD symptoms following ovarian suppression. Am J Psychiatry. 2017;174:980–9.
42. Dogra TD, Leenaars AA, Raintji R, Lalwani S, Girdhar S, Wenckstern S, et al. Menstruation and suicide: an exploratory study. Psychol Rep. 2007;101:430–4.
43. Zengin Y, Çalık I, Büyükcam F, Şen J, Akpınar Ş, Erdem A, et al. The relationship between suicide attempts and menstrual cycles in the emergency department and the sociodemographic and clinical characteristics of these patients. Eurasia J Emerg Med. 2015;14:118–22.
44. Afzali S, Taheri SK, Jamilian M, Eslambolchi P. The relationship between menstrual cycle phases and suicide attempts in suicidal women admitted to the poisoning ward of Farshchian Hospital, Hamedan, Iran. Iran J Toxicol. 2012;5:531–4.
45. Baca-Garcia E, Diaz-Sastre C, Ceverino A, Perez-Rodriguez MM, Navarro-Jimenez R, Lopez-Castroman J, et al. Suicide attempts among women during low estradiol/low progesterone states. J Psychiatr Res. 2010;44:209–14.
46. Kim B, Kang E-S, Fava M, Mischoulon D, Soskin D, Yu B-H, et al. Follicle-stimulating hormone (FSH), current suicidal ideation and attempt in female patients with major depressive disorder. Psychiatry Res. 2013;210:951–6.
47. Sun BZ, Kangarloo T, Adams JM, Sluss PM, Welt CK, Chandler DW, et al. Healthy post-menarchal adolescent girls demonstrate multi-level reproductive axis immaturity. J Clin Endocrinol Metab. 2018;104:613–23.
48. Chatterton RT, Mateo ET, Hou N, Rademaker AW, Acharya S, Jordan VC, et al. Characteristics of salivary profiles of oestradiol and progesterone in premenopausal women. J Endocrinol. 2005;186:77–84.
49. Gann PH, Giovanazzi S, Van Horn L, Branning A, Chatterton RT.Saliva as a medium for investigating intra- and inter-individual differences in sex hormone levels in premenopausal women.Cancer Epidemiol Biomark Prev. 2001;10:59–64.
50. Landgren B-M, Undén A-L, Diczfalusy E. Hormonal profile of the cycle in 68 normally menstruating women. Eur J Endocrinol. 1980;94:89–98.
51. Shirtcliff EA, Dahl RE, Pollak SD. Pubertal development: correspondence between hormonal and physical development. Child Dev. 2009;80:327–37.
52. Hambridge HL, Mumford SL, Mattison DR, Ye A, Pollack AZ, Bloom MS, et al. The influence of sporadic anovulation on hormone levels in ovulatory cycles. Hum Reprod Oxf Engl. 2013;28:1687–94.
53. Ellison PT, Panter-Brick C, Lipson SF, O'Rourke MT. The ecological context of human ovarian function. Hum Reprod Oxf Engl. 1993;8:2248–58.
54. Lobo RA, Granger L, Goebelsmann U, Mishell DR. Elevations in unbound serum estradiol as a possible mechanism for inappropriate gonadotropin secretion in women with PCO. J Clin Endocrinol Metab. 1981;52:156–8.
55. Cover KK, Maeng LY, Lebrón-Milad K, Milad MR. Mechanisms of estradiol in fear circuitry: implications for sex differences in psychopathology. Transl Psychiatry. 2014;4:e422.
56. Maner JK, Miller SL. Hormones and social monitoring: menstrual cycle shifts in progesterone underlie women's sensitivity to social information. Evol Hum Behav. 2014;35:9–16.
57. Sheppard PAS, Choleris E, Galea LAM. Structural plasticity of the hippocampus in response to estrogens in female rodents. Mol Brain. 2019;12:22.

58. Wang ACJ, Hara Y, Janssen WGM, Rapp PR, Morrison JH. Synaptic estrogen receptor-α levels in prefrontal cortex in female rhesus monkeys and their correlation with cognitive performance. J Neurosci. 2010;30:12770–6.
59. Engman J, Sundström Poromaa I, Moby L, Wikström J, Fredrikson M, Gingnell M. Hormonal cycle and contraceptive effects on amygdala and salience resting-state networks in women with previous affective side effects on the pill. Neuropsychopharmacology. 2018;43:555–63.
60. Engman J, Linnman C, Van Dijk KRA, Milad MR. Amygdala subnuclei resting-state functional connectivity sex and estrogen differences. Psychoneuroendocrinology. 2016;63:34–42.
61. Miedl SF, Wegerer M, Kerschbaum H, Blechert J, Wilhelm FH. Neural activity during traumatic film viewing is linked to endogenous estradiol and hormonal contraception. Psychoneuroendocrinology. 2018;87:20–6.
62. Protopopescu X, Butler T, Pan H, Root J, Altemus M, Polanecsky M, et al. Hippocampal structural changes across the menstrual cycle. Hippocampus. 2008;18:985–8.
63. Lisofsky N, Mårtensson J, Eckert A, Lindenberger U, Gallinat J, Kühn S. Hippocampal volume and functional connectivity changes during the female menstrual cycle. NeuroImage. 2015;118:154–62.
64. Pletzer B, Harris T-A, Scheuringer A, Hidalgo-Lopez E. The cycling brain: menstrual cycle related fluctuations in hippocampal and fronto-striatal activation and connectivity during cognitive tasks. Neuropsychopharmacology. 2019;44:1867–75.
65. Albert K, Pruessner J, Newhouse P. Estradiol levels modulate brain activity and negative responses to psychosocial stress across the menstrual cycle. Psychoneuroendocrinology. 2015; 59:14–24.
66. Arélin K, Mueller K, Barth C, Rekkas PV, Kratzsch J, Burmann I, et al. Progesterone mediates brain functional connectivity changes during the menstrual cycle—a pilot resting state MRI study. Front Neurosci [Internet]. 2015 Feb 23 [cited 2020 Apr 11];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4337344/
67. Pritschet L, Santander T, Layher E, Taylor CM, Yu S, Miller MB, et al. Functional reorganization of brain networks across the human menstrual cycle. bioRxiv. 2019 Dec 6;866913.
68. Auerbach RP, Pagliaccio D, Allison GO, Alqueza KL, Alonso MF. Neural correlates associated with suicide and nonsuicidal self-injury in youth. Biol Psychiatry. 2020;89:119–33.
69. Rybaczyk LA, Bashaw MJ, Pathak DR, Moody SM, Gilders RM, Holzschu DL. An overlooked connection: serotonergic mediation of estrogen-related physiology and pathology. BMC Women's Health. 2005;5:12.
70. Sullivan GM, Oquendo MA, Milak M, Miller JM, Burke A, Ogden RT, et al. Positron emission tomography quantification of serotonin(1 A) receptor binding in suicide attempters with major depressive disorder. JAMA Psychiatry. 2015;72:169–78.
71. Oquendo MA, Galfalvy H, Sullivan GM, Miller JM, Milak MM, Sublette ME, et al. Positron emission tomographic imaging of the serotonergic system and prediction of risk and lethality of future suicidal behavior. JAMA Psychiatry. 2016;73:1048–55.
72. Mann JJ The serotonergic system in mood disorders and suicidal behaviour. Philos Trans R Soc B Biol Sci [Internet]. 2013 [cited 2020 Apr 11];368. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3638390/
73. Steinberg LJ, Rubin-Falcone H, Galfalvy HC, Kaufman J, Miller JM, Sublette ME, et al. Cortisol stress response and in vivo PET imaging of human brain serotonin 1A receptor binding. Int J Neuropsychopharmacol. 2019;22:329–38.
74. Underwood MD, Kassir SA, Bakalian MJ, Galfalvy H, Dwork AJ, Mann JJ, et al. Serotonin receptors and suicide, major depression, alcohol use disorder and reported early life adversity. Transl Psychiatry. 2018;8:1–15.
75. Mellon SH, Griffin LD. Neurosteroids: biochemistry and clinical significance. Trends Endocrinol Metab TEM. 2002;13:35–43.
76. Giordano R, Picu A, Bonelli L, Balbo M, Berardelli R, Marinazzo E, et al. Hypothalamus-pituitary-adrenal axis evaluation in patients with hypothalamo-pituitary disorders: comparison of different provocative tests. Clin Endocrinol (Oxf). 2008;68:935–41.
77. Gunn JF, Lester D, McSwain S. Testing the warning signs of suicidal behavior among suicide ideators using the 2009 National survey on drug abuse and health. Int J Emerg Ment Health. 2011;13:147–54.
78. Skilbeck KJ, Johnston GAR, Hinton T. Stress and GABAA receptors. J Neurochem. 2010;112:1115–30.
79. Hu W, Zhang M, Czéh B, Flügge G, Zhang W. Stress impairs GABAergic network function in the hippocampus by activating nongenomic glucocorticoid receptors and affecting the integrity of the parvalbumin-expressing neuronal network. Neuropsychopharmacology 2010;35:1693–707.
80. Zhao J, Verwer RWH, Gao S-F, Qi X-R, Lucassen PJ, Kessels HW, et al. Prefrontal alterations in GABAergic and glutamatergic gene expression in relation to depression and suicide. J Psychiatr Res. 2018;102:261–74.
81. Archer J. Testosterone and human aggression: an evaluation of the challenge hypothesis. Neurosci Biobehav Rev. 2006;30: 319–45.
82. van Anders SM. Beyond masculinity: testosterone, gender/sex, and human social behavior in a comparative context. Front Neuroendocrinol. 2013;34:198–210.
83. Halpern CT, Udry JR, Campbell B, Suchindran C. Relationships between aggression and pubertal increases in testosterone: a panel analysis of adolescent males. Soc Biol. 1993;40:8–24.
84. Rice TR, Sher L Adolescent suicide and testosterone: Postnatal testosterone may be an important mediator of the association between prematurity and male neurodevelopmental disorders: A Hypothesis. Int J Adolesc Med Health [Internet]. 2015 Nov 4 [cited 2020 Apr 11];29. Available from: https://www.degruyter.com/view/journals/ijamh/29/4/article-20150058.xml
85. Perez-Rodriguez MM, Lopez-Castroman J, Martinez-Vigo M, Diaz-Sastre C, Ceverino A, Núñez-Beltrán A, et al. Lack of association between testosterone and suicide attempts. Neuropsychobiology. 2011;63:125–30.
86. McGirr A, Turecki G. The relationship of impulsive aggressiveness to suicidality and other depression-linked behaviors. Curr Psychiatry Rep. 2007;9:460–6.
87. Eisenegger C, Haushofer J, Fehr E. The role of testosterone in social interaction. Trends Cogn Sci. 2011;15:263–71.
88. Duke SA, Balzer BWR, Steinbeck KS. Testosterone and its effects on human male adolescent mood and behavior: a systematic review. J Adolesc Health. 2014;55:315–22.
89. van Bokhoven I, van Goozen SHM, van Engeland H, Schaal B, Arseneault L, Séguin JR, et al. Salivary testosterone and aggression, delinquency, and social dominance in a population-based longitudinal study of adolescent males. Horm Behav. 2006;50:118–25.
90. Schaal B, Tremblay RE, Soussignan R, Susman EJ. Male testosterone linked to high social dominance but low physical aggression in early adolescence. J Am Acad Child Adolesc Psychiatry. 1996;35:1322–30.
91. Milner A, Page A, Morrell S, Hobbs C, Carter G, Dudley M, et al. The effects of involuntary job loss on suicide and suicide attempts among young adults: evidence from a matched case-control study. Aust N Z J Psychiatry. 2014;48:333–40.
92. Stefansson J, Chatzittofis A, Nordström P, Arver S, Åsberg M, Jokinen J. CSF and plasma testosterone in attempted suicide. Psychoneuroendocrinology 2016;74:1–6.

93. Gustavsson G, Träskman-Bendz L, Higley JD, Westrin Å. CSF testosterone in 43 male suicide attempters. Eur Neuropsychopharmacol. 2003;13:105–9.
94. Zhang J, Jia C-X, Wang L-L. Testosterone differs between suicide attempters and community controls in men and women of China. Physiol Behav. 2015;141:40–5.
95. Roland BC, Morris JL, Zelhart PF. Proposed relation of testosterone levels to male suicides and sudden deaths. Psychol Rep. 1986;59:100–2.
96. Sher L, Grunebaum MF, Sullivan GM, Burke AK, Cooper TB, Mann JJ, et al. Testosterone levels in suicide attempters with bipolar disorder. J Psychiatr Res. 2012;46:1267–71.
97. Sher L, Grunebaum MF, Sullivan GM, Burke AK, Cooper TB, Mann JJ, et al. Association of testosterone levels and future suicide attempts in females with bipolar disorder. J Affect Disord. 2014;166:98–102.
98. Tripodianakis J, Markianos M, Rouvali O, Istikoglou C. Gonadal axis hormones in psychiatric male patients after a suicide attempt. Eur Arch Psychiatry Clin Neurosci. 2007;257:135–9.
99. Markianos M, Tripodianakis J, Istikoglou C, Rouvali O, Christopoulos M, Papageorgopoulos P, et al. Suicide attempt by jumping: a study of gonadal axis hormones in male suicide attempters versus men who fell by accident. Psychiatry Res. 2009;170:82–5.
100. Shay DA, Vieira-Potter VJ, Rosenfeld CS Sexually Dimorphic Effects of Aromatase on Neurobehavioral Responses. Front Mol Neurosci [Internet]. 2018 [cited 2020 Apr 17];11. Available from: https://www.frontiersin.org/articles/10.3389/fnmol.2018.00374/full
101. Fitzgerald ML, Kassir SA, Underwood MD, Bakalian MJ, Mann JJ, Arango V. Dysregulation of Striatal Dopamine Receptor Binding in Suicide. Neuropsychopharmacology. 2017;42:974–82.
102. Pitchot W, Hansenne M, Ansseau M. Role of dopamine in non-depressed patients with a history of suicide attempts. Eur Psychiatry. 2001;16:424–7.
103. Roy A, Karoum F, Pollack S. Marked reduction in indexes of dopamine metabolism among patients with depression who attempt suicide. Arch Gen Psychiatry. 1992;49:447–50.
104. Tobiansky DJ, Wallin-Miller KG, Floresco SB, Wood RI, Soma KK Androgen Regulation of the Mesocorticolimbic System and Executive Function. Front Endocrinol [Internet]. 2018 [cited 2020 Apr 14];9. Available from: https://www.frontiersin.org/articles/10.3389/fendo.2018.00279/full
105. Sinclair D, Purves-Tyson TD, Allen KM, Weickert CS. Impacts of stress and sex hormones on dopamine neurotransmission in the adolescent brain. Psychopharmacol. 2014;231:1581–99.
106. Wahlstrom D, Collins P, White T, Luciana M. Developmental changes in dopamine neurotransmission in adolescence: behavioral implications and issues in assessment. Brain Cogn. 2010;72:146–59.
107. Pitchot W, Hansenne M, Moreno AG, Ansseau M. Dopamine and history of suicide attempts. Clin Neuropharmacol. 1992;15:296B.
108. Happe S, Tings T, Koch W, Welsch J, Helmschmied K, Baier PC, et al. Growth hormone response in low-dose apomorphine test correlates with nigrostriatal dopamine transporter binding in patients with Parkinson's disease. J Neural Transm Vienna Austria 1996. 2007;114:589–94.
109. Oquendo MA, Sullivan GM, Sudol K, Baca-Garcia E, Stanley BH, Sublette ME, et al. Toward a Biosignature for Suicide. Am J Psychiatry. 2014;171:1259–77.
110. Barendse MEA, Simmons JG, Byrne ML, Seal ML, Patton G, Mundy L, et al. Brain structural connectivity during adrenarche: associations between hormone levels and white matter microstructure. Psychoneuroendocrinology. 2018;88:70–7.
111. Menzies L, Goddings A-L, Whitaker KJ, Blakemore S-J, Viner RM. The effects of puberty on white matter development in boys. Dev Cogn Neurosci. 2015;11:116–28.
112. Peper JS, Reus MA, de, Heuvel MP, van den, Schutter DJLG. Short fused? associations between white matter connections, sex steroids, and aggression across adolescence. Hum Brain Mapp. 2015;36:1043–52.
113. Bos PA, van Honk J, Ramsey NF, Stein DJ, Hermans EJ. Testosterone administration in women increases amygdala responses to fearful and happy faces. Psychoneuroendocrinology. 2013;38:808–17.
114. Spielberg JM, Olino TM, Forbes EE, Dahl RE. Exciting fear in adolescence: does pubertal development alter threat processing? Dev Cogn Neurosci. 2014;8:86–95.
115. Cservenka A, Stroup ML, Etkin A, Nagel BJ. The effects of age, sex, and hormones on emotional conflict-related brain response during adolescence. Brain Cogn. 2015;99:135–50.
116. Op de Macks ZA, Moor BG, Overgaauw S, Güroğlu B, Dahl RE, Crone EA. Testosterone levels correspond with increased ventral striatum activation in response to monetary rewards in adolescents. Dev Cogn Neurosci. 2011;1:506–16.
117. Laube C, van den Bos W, Fandakova Y. The relationship between pubertal hormones and brain plasticity: Implications for cognitive training in adolescence. Dev Cogn Neurosci. 2020;42:100753.
118. Braams BR, van Duijvenvoorde ACK, Peper JS, Crone EA. Longitudinal changes in adolescent risk-taking: a comprehensive study of neural responses to rewards, pubertal development, and risk-taking behavior. J Neurosci J Soc Neurosci. 2015;35:7226–38.
119. White SF, Lee Y, Schlund MW, Shirtcliff EA, Ladouceur CD. Testosterone reactivity is associated with reduced neural response to reward in early adolescence. Behav Brain Res. 2020;387:112593.
120. Coates JM, Herbert J. Endogenous steroids and financial risk taking on a London trading floor. Proc Natl Acad Sci. 2008;105:6167–72.
121. Stanton SJ, Liening SH, Schultheiss OC. Testosterone is positively associated with risk taking in the Iowa Gambling Task. Horm Behav. 2011;59:252–6.
122. Richard-Devantoy S, Berlim MT, Jollant F. A meta-analysis of neuropsychological markers of vulnerability to suicidal behavior in mood disorders. Psychol Med. 2014;44:1663–73.
123. Gifuni AJ, Perret LC, Lacourse E, Geoffroy M-C, Mbekou V, Jollant F, et al. Decision-making and cognitive control in adolescent suicidal behaviors: a qualitative systematic review of the literature. Eur Child Adolesc Psychiatry [Internet]. 2020 May 9 [cited 2021 Mar 8]; Available from: https://doi.org/10.1007/s00787-020-01550-3
124. Chatzittofis A, Nordström P, Hellström C, Arver S, Åsberg M, Jokinen J. CSF 5-HIAA, cortisol and DHEAS levels in suicide attempters. Eur Neuropsychopharmacol. 2013;23:1280–7.
125. Butterfield MI, Stechuchak KM, Connor KM, Davidson JRT, Wang C, MacKuen CL, et al. Neuroactive steroids and suicidality in posttraumatic stress disorder. Am J Psychiatry. 2005;162:380–2.
126. Sher L, Flory J, Bierer L, Makotkine I, Yehuda R. Dehydroepiandrosterone and dehydroepiandrosterone sulfate levels in combat veterans with or without a history of suicide attempt. Acta Psychiatr Scand. 2018;138:55–61.
127. Sequeira A, Shen K, Gottlieb A, Limon A.Human brain transcriptome analysis finds region- and subject-specific expression signatures of GABA A R subunits. Commun Biol. 2019;2:1–14.
128. Pineles SL, Nillni YI, Pinna G, Irvine J, Webb A, Arditte Hall KA, et al. PTSD in women is associated with a block in conversion of progesterone to the GABAergic neurosteroids allopregnanolone and pregnanolone measured in plasma. Psychoneuroendocrinology. 2018;93:133–41.
129. Locci A, Pinna G. Neurosteroid biosynthesis down-regulation and changes in GABAA receptor subunit composition: a biomarker axis in stress-induced cognitive and emotional impairment. Br J Pharm. 2017;174:3226–41.

130. Ellis R, Fernandes A, Simmons JG, Mundy L, Patton G, Allen NB, et al. Relationships between adrenarcheal hormones, hippocampal volumes and depressive symptoms in children. Psychoneuroendocrinology. 2019;104:55–63.
131. Barendse MEA, Simmons JG, Patton G, Mundy L, Byrne ML, Seal ML, et al. Adrenarcheal timing longitudinally predicts anxiety symptoms via amygdala connectivity during emotion processing. J Am Acad Child Adolesc Psychiatry [Internet]. 2019 May 2 [cited 2020 Apr 12]. Available from: https://www.jaacap.org/article/S0890-8567(19)30286-2/abstract
132. Whittle S, Simmons JG, Byrne ML, Strikwerda-Brown C, Kerestes R, Seal ML, et al. Associations between early adrenarche, affective brain function and mental health in children. Soc Cogn Affect Neurosci. 2015;10:1282–90.
133. Sripada RK, Marx CE, King AP, Rajaram N, Garfinkel SN, Abelson JL, et al. DHEA enhances emotion regulation neurocircuits and modulates memory for emotional stimuli. Neuropsychopharmacology. 2013;38:1798–807.
134. McLoughlin AB, Gould MS, Malone KM. Global trends in teenage suicide: 2003–2014. QJM Mon J Assoc Phys. 2015;108:765–80.
135. WHO | Preventing suicide: a global imperative [Internet]. WHO. World Health Organization; [cited 2020 Mar 26]. Available from: http://www.who.int/mental_health/suicide-prevention/world_report_2014/en/
136. Ruch DA, Sheftall AH, Schlagbaum P, Rausch J, Campo JV, Bridge JA.Trends in suicide among youth aged 10 to 19 years in the United States, 1975 to 2016. JAMA Netw Open.2019;2:e193886.
137. Lewinsohn PM, Rohde P, Seeley JR, Baldwin CL. Gender differences in suicide attempts from adolescence to young adulthood. J Am Acad Child Adolesc Psychiatry. 2001;40:427–34.
138. Boeninger DK, Masyn KE, Feldman BJ, Conger RD. Sex differences in developmental trends of suicide ideation, plans, and attempts among European American adolescents. Suicide Life Threat Behav. 2010;40:451–64.
139. Miranda-Mendizabal A, Castellví P, Parés-Badell O, Alayo I, Almenara J, Alonso I, et al. Gender differences in suicidal behavior in adolescents and young adults: systematic review and meta-analysis of longitudinal studies. Int J Public Health. 2019;64:265–83.
140. Brent DA, Baugher M, Bridge J, Chen T, Chiappetta L. Age- and sex-related risk factors for adolescent suicide. J Am Acad Child Adolesc Psychiatry. 1999;38:1497–505.
141. Dorn LD, Biro FM. Puberty and its measurement: a decade in review. J Res Adolesc. 2011;21:180–95.
142. Graber JA. Pubertal timing and the development of psychopathology in adolescence and beyond. Horm Behav. 2013;64:262–9.
143. Kaltiala-Heino R, Marttunen M, Rantanen P, Rimpelä M. Early puberty is associated with mental health problems in middle adolescence. Soc Sci Med. 2003;57:1055–64.
144. Lee D, Ahn I-Y, Park C-S, Kim B-J, Lee C-S, Cha B, et al. Early menarche as a risk factor for suicidal ideation in girls: The Korea youth risk behavior web-based survey. Psychiatry Res. 2020;285:112706.
145. Nacinovich R, Buzi F, Oggiano S, Rossi S, Spada S, Broggi F, et al. Body experiences and psychopathology in idiopathic central precocious and early puberty. Minerva Pediatr. 2016;68:11–8.
146. Stice E, Presnell K, Bearman SK. Relation of early menarche to depression, eating disorders, substance abuse, and comorbid psychopathology among adolescent girls. Dev Psychol. 2001;37:608–19.
147. Roberts E, Fraser A, Gunnell D, Joinson C, Mars B. Timing of menarche and self-harm in adolescence and adulthood: a population-based cohort study. Psychol Med. 2020;50:2010–8.
148. Belsky J, Steinberg LD, Houts RM, Friedman SL, DeHart G, Cauffman E, et al. Family rearing antecedents of pubertal timing. Child Dev. 2007;78:1302–21.
149. Turecki G, Ernst C, Jollant F, Labonté B, Mechawar N. The neurodevelopmental origins of suicidal behavior. Trends Neurosci. 2012;35:14–23.
150. Wichstrøm L, Hegna K. Sexual orientation and suicide attempt: A longitudinal study of the general Norwegian adolescent population. J Abnorm Psychol. 2003;112:144–51.
151. Klonsky ED, Saffer BY, Bryan CJ. Ideation-to-action theories of suicide: a conceptual and empirical update. Curr Opin Psychol. 2018;22:38–43.
152. Beauchaine TP, Crowell SE, Hsiao RC. Post-dexamethasone cortisol, self-inflicted injury, and suicidal ideation among depressed adolescent girls. J Abnorm Child Psychol. 2015;43:619–32.
153. Shalev A, Porta G, Biernesser C, Zelazny J, Walker-Payne M, Melhem N, et al. Cortisol response to stress as a predictor for suicidal ideation in youth. J Affect Disord. 2019;257:10–6.
154. King LS, Graber MG, Colich NL, Gotlib IH. Associations of waking cortisol with DHEA and testosterone across the pubertal transition: effects of threat-related early life stress. Psychoneuroendocrinology. 2020;115:104651.
155. Maninger N, Wolkowitz OM, Reus VI, Epel ES, Mellon SH. Neurobiological and neuropsychiatric effects of dehydroepiandrosterone (DHEA) and DHEA sulfate (DHEAS). Front Neuroendocrinol. 2009;30:65–91.
156. Oyola MG, Handa RJ. Hypothalamic-pituitary-adrenal and hypothalamic-pituitary-gonadal axes: sex differences in regulation of stress responsivity. Stress Amst Neth 2017;20:476–94.
157. Toufexis D, Rivarola MA, Lara H, Viau V. Stress and the reproductive axis. J Neuroendocrinol. 2014;26:573–86.
158. Mastorakos G, Pavlatou MG, Mizamtsidi M. The hypothalamic-pituitary-adrenal and the hypothalamic- pituitary-gonadal axes interplay. Pediatr Endocrinol Rev. 2006;3:172–81.
159. Mendle J, Beltz AM, Carter R, Dorn LD. Understanding puberty and its measurement: ideas for research in a new generation. J Res Adolesc. 2019;29:82–95.
160. Kleiman EM, Turner BJ, Fedor S, Beale EE, Picard RW, Huffman JC, et al. Digital phenotyping of suicidal thoughts. Depress Anxiety. 2018;35:601–8.
161. Rosenfield RL. Adolescent anovulation: maturational mechanisms and implications. J Clin Endocrinol Metab. 2013;98:3572–83.
162. Østby Y, Tamnes CK, Fjell AM, Westlye LT, Due-Tønnessen P, Walhovd KB. Heterogeneity in subcortical brain development: a structural magnetic resonance imaging study of brain maturation from 8 to 30 years. J Neurosci. 2009;29:11772–82.



Review in Advance first posted online on September 10, 2015. (Changes may still occur before final publication online and in print.)

Psychology of Habit

Wendy Wood and Dennis Rünger

Department of Psychology, University of Southern California, Los Angeles, California 90089-1061; email: wendy.wood@usc.edu, dennis.ruenger@gmail.com

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

Annu. Rev. Psychol. 2016. 67:11.1–11.26

The *Annual Review of Psychology* is online at psych.annualreviews.org

This article's doi: 10.1146/annurev-psych-122414-033417

Copyright © 2016 by Annual Reviews. All rights reserved

Keywords

automaticity, dual-process models, goals, behavior change, model-free learning

Abstract

As the proverbial creatures of habit, people tend to repeat the same behaviors in recurring contexts. This review characterizes habits in terms of their cognitive, motivational, and neurobiological properties. In so doing, we identify three ways that habits interface with deliberate goal pursuit: First, habits form as people pursue goals by repeating the same responses in a given context. Second, as outlined in computational models, habits and deliberate goal pursuit guide actions synergistically, although habits are the efficient, default mode of response. Third, people tend to infer from the frequency of habit performance that the behavior must have been intended. We conclude by applying insights from habit research to understand stress and addiction as well as the design of effective interventions to change health and consumer behaviors.

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

Contents

INTRODUCTION 11.2
HABIT AUTOMATICITY 11.4
Automatic Cuing of Habits 11.4
Habit Automaticity and Deliberation About Action 11.5
HABIT FORMATION 11.6
Associative and Reward Mechanisms in Habit Learning 11.6
Measures of Habit Strength 11.7
COMPUTATIONAL MODELS 11.8
NEUROBIOLOGY OF HABITS 11.10
Neural Systems Associated with Habits 11.10
Neural Systems that Integrate Habit and Goal-Directed Action Control 11.12
FACTORS THAT SHIFT THE BALANCE BETWEEN HABITS AND DELIBERATE GOAL PURSUIT 11.12
Stress 11.13
Addiction 11.14
INFERENCES ABOUT THE CAUSES OF HABIT PERFORMANCE 11.15
CHANGING HABITS 11.16
Interventions to Impede Unwanted Habit Performance 11.17
Interventions to Promote Formation of Desired Habits 11.17
CONCLUSION 11.18

INTRODUCTION

The yin and yang of the history of habits is closely tied to broader trends in the history of psychology. William James's (1890) view that "habit covers a very large part of life," necessitated that we "define clearly just what its limits are" (p. 104). Psychology complied, and the habit construct acquired specific meanings in the behaviorist traditions of Thorndike's (1898) law of effect, Hull's (1943) formalized drive theory, and Skinner's (1938) operant conditioning. However, these reinforcement-based models of habit were soon supplanted as the field embraced more purposive and cognitive perspectives. To Tolman (1948), repeated behaviors reflected learning of internal representations and maps, and Miller et al. (1960) argued that habits be replaced with information-processing mechanisms of goal pursuit. Accordingly, cognitive psychology and decision-making research in the 1960s and 70s developed largely separately from research on habit.

More recent theories have captured the complexity of action control and enabled integration of these opposing conceptualizations. The concept of automaticity (Shiffrin & Schneider 1977) and theories of dual information processing (Wason & Evans 1975) provided frameworks inclusive of habits and thoughtful decision making. Through procedural memory, habits could be cognitively represented as distinct from other types of implicit processes as well as from explicit, declarative memories (Squire & Zola-Morgan 1991). Reinforcement learning (RL) research with animals identified a behavioral criterion for detecting habit performance, involving insensitivity to changes in rewarding outcomes (Dickinson 1985), and neuroscience began to identify the brain regions and circuits involved in habitual behavior (Graybiel 1998). Social-cognitive approaches outlined a variety of ways in which habits interface with goals (Verplanken & Aarts 1999, Ouellette & Wood 1998), and computational models included goal pursuit and prospective planning as

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

well as habit-like mechanisms of stimulus-driven or model-free action control (Cooper et al. 2014, Daw et al. 2005). Empirical fuel for the study of habit was provided by evidence of the high levels of repetition in daily activities (Khare & Inman 2006, Wood et al. 2002). Further framing these developments is the reading public's interest in understanding their own habits (e.g., Rubin 2015).

In this article, we take stock of the fast-growing research on habits. Our review is necessarily wide reaching, covering a variety of behavioral domains, cognitive tasks, and neuroscientific findings. We also consider animal learning research when relevant, given that habit learning mechanisms are largely conserved across mammalian species and do not appear to be degraded in humans or replaced by higher cortical functions (Bayley et al. 2005). Across these diverse research paradigms, the common elements are habit learning through repeated responding so as to form context-response associations in memory, and automated habit performance that is relatively insensitive to changes in the value or contingency of response outcomes. As we explain, these definitions of habit learning and performance capture the cognitive and neural mechanisms involved in habit memory and are reflected in characteristic patterns in habitual responding.

The schematic in **Figure 1** provides a framework for this review by depicting three ways in which habits interface with goals to guide behavior. Goals energize and direct action by defining a desired end state. In our three-pronged model, habits and goals interact through habit formation, habit performance, and inferences about the causes of behavior. First, goals influence habit formation by initially motivating people to repeat actions and to expose themselves to performance contexts (see Habit Formation section). This is illustrated by the arrows from goal system to context cues and habitual response in **Figure 1**. Once habits form, context cues come to automatically activate the habit representation in memory (see Habit Automaticity section). Second, people act on the habit in mind as well as on their prevailing goals by tailoring their behavior to the current circumstances (see Computational Models section). External factors such as stress and distraction influence the impact of these two processes by reducing the motivation or ability to deliberately pursue goals and increasing reliance on habits (see section Factors that Shift the Balance Between Habits and Goals section). As outlined in dual-process frameworks, habits provide a sort of default response unless people are sufficiently motivated and able to tailor their behavior to current circumstances. Finally, people make inferences about their goals based on observing their own frequent behavior, as reflected by the double-headed arrow in **Figure 1** between the habitual response and the goal system (see Inferences about the Causes of Habit Performance section).

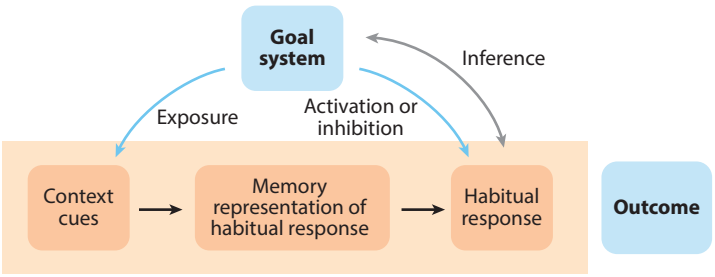


Figure 1

Schematic of three ways in which habits interface with deliberate goal pursuit: through initial repetition and exposure to contexts during habit formation (illustrated by the arrows from goal system to context cues and habitual response), through activation or inhibition of the habitual response, and through inferences about the probable causes of habit responding (reflected by the double-headed arrow between habitual response and goal system).

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

To conclude the review, we apply insights from basic research on habit to understand stress and addiction and to design successful behavior-change interventions (see Changing Habits section).

HABIT AUTOMATICITY

The terms habit and automaticity are sometimes used interchangeably. Like other automatic responses, habits are activated in memory in an autonomous fashion without requiring executive control (Evans & Stanovich 2013). Habits, however, are not synonymous with automaticity but are best understood as learned automatic responses with specific features (Wood et al. 2014). Two defining features of habit automaticity are (*a*) activation by recurring context cues and (*b*) insensitivity to short-term changes in goals (a.k.a., not goal dependent), including changes in the value of response outcomes and the response-outcome contingency. Additional features that apply to most habitual responses include speed and efficiency, limited thought, rigidity, and integration of sequences of responses that can be executed as a unit (Seger & Spiering 2011, Smith & Graybiel 2013). However, each of these additional features may not be assessed in all habit research paradigms. Instrumental learning studies, for example, often do not evaluate response speed, whereas learning of motor sequences in cognitive-experimental research is assessed primarily via changes in response latency.

Habits differ from other automatic, implicit processes including priming, classical conditioning, and nonassociative learning (Evans & Stanovich 2013, Squire & Zola-Morgan 1991). For example, the priming of goals, attitudes, or concepts can activate a range of responses, not only the repetition of a particular well-learned response (see Wood et al. 2014). Even strongly desired goals that stably characterize people's motives do not necessarily yield stability in the particular means of goal pursuit. In contrast, habit automaticity applies to a specific response. Furthermore, unlike habits, automated goals (e.g., implementation intentions) influence behavior primarily to the extent that they are consistent with people's explicit motivations (Sheeran et al. 2005).

Automatic Cuing of Habits

A variety of cues might trigger habit performance, including aspects of physical environments, other people, and preceding actions in a sequence. Once habits form, perception of the relevant context cues automatically activates the mental representation of the habitual response. Exposure to cues might be deliberate, as when sitting at a computer in order to activate thoughts of work. Or exposure can be inadvertent, as when a chance sighting of a fast-food outlet activates thoughts of eating. We assume that the memory representation of a habit response is cognitively richer than a mere motor program that controls response execution. Given that human cognition is based on integrated sensorimotor units (Hommel 2009), a habitual response will be represented in terms of response features as well as perceptual features. Specifically, the sensory feedback while making a response, which gives rise to the experience of performing the action, is included in the mental representation. As a consequence, a habit cue not only triggers a motor program, it also activates a multimodal representation, or thought, of the habitual response. Consistent with this view, Neal et al. (2012) found that runners with strong running habits automatically brought to mind thoughts of running and jogging when exposed to words designating the physical locations in which they typically ran.

Once habitual responses are activated, people can act on the response in mind without making a decision to do so. That is, habit performance follows relatively directly from the perception of context cues and thoughts about the behavior, reflecting the tight linkage between an internal action representation and the action itself (James 1890). For example, when students who

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

frequently went to the sports stadium on campus were incidentally exposed to an image of the stadium, they raised their voices as they would habitually in that context, despite no change in their motivation to speak loudly (Neal et al. 2012). Also, in a study conducted in a local cinema, participants with stronger habits to eat popcorn at the movies consumed more than those with weak habits, even when they disliked the popcorn because it was stale and unpalatable (Neal et al. 2011). However, when in a campus meeting room watching music videos, participants with strong cinema-popcorn eating habits were guided by their preferences and ate little stale popcorn.

Habit performance is typified by the insensitivity to outcomes apparent in speaking loudly in a quiet laboratory setting and eating popcorn despite disliking it. This insensitivity has been demonstrated directly in instrumental learning experiments in which participants were first trained extensively to choose a reward to a certain image cue (e.g., Tricomi et al. 2009). Participants then ate as much of the reward as they desired, so that they did not want any more of that specific food. Nonetheless, when tested again, extensively trained participants in this paradigm continued to make the habitual but unwanted choice to the associated image (see also Hogarth et al. 2012b, Schwabe & Wolf 2010).

Triandis's (1977) early work on behavior prediction sparked social psychologists' interest in the idea that repeated behavior becomes more habitual and less dependent on goal pursuit. In prediction studies, behavioral intentions and habit strength (usually operationalized as frequency of past performance) are used to predict future performance. In a meta-analysis of 64 such studies, Ouellette & Wood (1998) found that intentions were stronger predictors of actions that were performed only occasionally (e.g., getting flu shots) than actions that could be repeated more regularly (e.g., wearing seat belts). Actions performed regularly apparently became habitual and persisted with little guidance from intentions (see also Gardner et al. 2011). In addition, habits and intentions interact in guiding daily variations in behavior. For example, on days when participants' intentions to engage in physical activity were weaker than usual, they fell back on their exercise habits and worked out only to the extent that exercise was habitual (Rebar et al. 2014). Also, in a longitudinal test, habit strength to donate blood determined the relation between intentions and actual donations (P. Sheeran, G. Godin, M. Conner, and M. Germain, unpublished data). That is, for participants with weak donation habits, increasingly favorable intentions predicted a greater number of future donations. However, as habit strength increased, the predictive power of intentions diminished, and participants with the strongest habits simply repeated their past donations without input from intentions.

When acting out of habit, the ready response in mind reduces deliberation and narrows focus even when some explicit decision making is required. In a multiattribute choice task involving a series of travel mode decisions, participants with stronger habits to ride a bike or drive their car conducted less extensive information searches, considered fewer action alternatives, and biased the searches toward their habitual choice (Aarts et al. 1997; see also Betsch et al. 2001). Experimentally enhancing attention to the decision process temporarily increased alternative choices, but the habitual choice reemerged with continued decisions (Verplanken et al. 1997). These results may in part reflect that the repeated activation of one response in a context reduces the cognitive accessibility of alternatives (Danner et al. 2007). Essentially, people with strong habits process information in ways that reduce the likelihood that they will consider acting otherwise.

Habit Automaticity and Deliberation About Action

Dual-process theories of decision making and judgment outline the mechanisms that lead people to respond automatically or to engage in deliberate information processing that draws on the limited-capacity resource of working memory (Evans & Stanovich 2013). From this view, people

REVIEWS IN ADVANCE
Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

often act habitually in the interests of efficiency (Wood et al. 2014). When motivated and able to engage in deliberate goal pursuit, however, they might identify desired outcomes, set and initiate behavioral intentions, end actions, and evaluate outcomes (Gollwitzer & Brandstätter 1997).

Dual-process models in psychology often specify that automatic and deliberate systems interact through a default-interventionist architecture (Evans & Stanovich 2013) so that responses are largely habitual unless the deliberative system intervenes to impose an alternative. In contrast, reinforcement-learning computational models of routine behavior and decision making (see Computational Models section) often invoke the less psychologically plausible parallel-competitive form of dual-process architecture, in which planning proceeds in parallel with habitual control. As Evans & Stanovich (2013) point out, parallel processing assumes that a costly central executive is almost continuously ready to be engaged in planning and deliberation. More recently, some computational models have adapted a more default interventionist approach in which planning intervenes to alter habits only when necessary (e.g., Pezzulo et al. 2013). It is useful to note that, although we invoke the dual-process framework in this review, action control is influenced by more than two processes, including automated goal pursuit (Sheeran et al. 2005) as well as Pavlovian conditioning of incentive motivation (Balleine & O'Doherty 2010).

In summary, perception of habit cues automatically activates habit representations, and people typically carry out the habit in mind. Thus, behavior prediction studies reveal that people often act out of habit, even when it is in conflict with their intentions. In dual-process models depicting the ways that habits integrate with goals in guiding behavior, habits are a ready default unless people are motivated and able to intervene and engage in more deliberate goal pursuit.

HABIT FORMATION

Given that everyday habits develop as people go about pursuing life goals, habit formation is closely intertwined with goal pursuit. Nonetheless, an implication of the basic context-response mechanism underlying habits is that behavior becomes less responsive to current goals and planning as habit associations strengthen.

Habits develop through instrumental learning and build on the fundamental principle that rewarded responses are repeated (Thorndike 1898). When repeatedly pursuing a goal such as making coffee, people experience covariations between context cues (e.g., coffee filter) and actions (e.g., measure grounds) that lead to goal attainment. Daily life is full of such repetition. In experience-sampling research in which people recorded once per hour what they were thinking, feeling, and doing, about 43% of actions were performed almost daily and usually in the same context (Wood et al. 2002; see also Khare & Inman 2006). Particular actions, such as types of food eaten, also tend to be performed in particular physical locations (Liu et al. 2015). Typically, the learning of context-response associations is an unintended consequence of this repetition. Suggestive of this automaticity, participants in Wood et al.'s (2002) study often reported that they did not think about repeated behaviors during performance.

Associative and Reward Mechanisms in Habit Learning

Habits strengthen through associative and reward-learning mechanisms that capture the slow, incremental nature of habit formation. With each repetition, small changes occur in the cognitive and neural mechanisms associated with procedural memory. Through Hebbian learning of repeated connections, cognitive associations between context cues and a response are strengthened gradually so that people are prepared to repeat performance when the context cues are encountered again (called direct cuing by Wood & Neal 2007).

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

The strength of context-response associations is further modulated by the reward following the response. At a neural level, midbrain dopamine systems support this reinforcement process. By signaling reward prediction errors, or the discrepancy between an anticipated and actual reward, a phasic dopamine response acts as a teaching signal for habit learning in the striatum (Balleine & O'Doherty 2010). Specifically, the dopaminergic signal that is triggered by an unexpected change in reward magnitude works retroactively to stamp in associations between the still-active memory traces of the response and the cues in the performance context (Wise 2004). Thus, dopamine signals promote habit learning as people initially repeat responses to a reward, but the signals become less active with repetition, as the reward recurs.

In computational RL models, habit formation is conceptualized as the learning of value signals that represent the expected future rewards for the different response options in a given context and provide the basis for selecting an action (see Computational Models section). Value-based selection of habitual responses can be can be regarded as a form of motivated cuing (Wood & Neal 2007).

Another way in which habitual responding continues to be influenced by motivational processes is through context cues that have become associated with the reward that follows an action. As learned predictors of reward, such Pavlovian context cues can cause the habitual response to be performed with increased vigor (Balleine & O'Doherty 2010). Of note, the motivational effect of such Pavlovian predictors of rewarding outcomes is distinct from the motivational value of the outcome itself. Holland (2004), for example, found that responses extensively trained into habits were insensitive to changes in outcome value but continued to be influenced by reward-related cues.

A standard finding from animal learning is that habits develop most readily when rewards are provided on interval schedules, meaning that responses are rewarded only after a time has elapsed. Such rewards mimic natural resources that are replenished over time. With these schedules, changes in response rate during the time interval do not change the amount of reward delivered, reducing the experience of instrumental contingency between the response and the reward (Dickinson 1985). Thus, interval rewards likely promote habit learning because context-response associations can form without including a representation of the goal or outcome of the action.

As a caveat to the principle that habits form from repetition, habits do not always emerge with complex tasks in which different response choices lead to different rewards. In animal research, even after extensive training at a task involving a lever press (yielding sucrose) or chain pull (yielding a food pellet), rats failed to form habits—they continued to be sensitive to reward value and ceased responding for one of these rewards after it had been paired with a toxin (Colwill & Rescorla 1985). Perhaps in an analogous fashion, decision making in humans impedes habit formation. In a repeated sequential choice task, participants failed to form habits to the extent that they spontaneously used a planning strategy and based their choices on the value and probability of response outcomes (Gillan et al. 2015). Similarly, habit formation was hindered when an instrumental task promoted planning compared with mapping of responses to cues (Liljeholm et al. 2015). It appears, then, that deliberative decision making is protective against habit formation even when people respond repeatedly to particular cues.

In summary, habits are likely to form from responses repeated contiguously with context cues, especially when responses are rewarded on an interval schedule. Through the activation of dopamine systems, habits form that are insensitive to current shifts in reward value and structure. However, planning and making deliberate choices during responding can hinder habit formation.

Measures of Habit Strength

Habitization is a process, with no clear demarcation point when strong habits have formed. Nonetheless, instrumental learning tasks provide a clear behavioral criterion. That is, habits have

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

formed when participants continue to repeat a well-practiced response after the reward has been reduced in value (e.g., for food rewards, after consuming to satiety) or is no longer tied to the response (Dickinson 1985). In these paradigms, habitual responding is evaluated during a subsequent extinction phase (rewards withheld) in order to preclude additional learning based on the changed reward values.

Behavioral indicators of habit strength are captured in a variety of experimental paradigms beyond the simple motor responses often thought emblematic of habits. Habit strength has been manipulated in experimental paradigms involving word-association tasks (Hay & Jacoby 1996, Quinn et al. 2010), choice tasks involving pictorial and other judgment stimuli (de Wit et al. 2009, Gillan et al. 2015), two-stage decision-making tasks (Daw et al. 2011), and problem-solving tasks such as tower building (Patsenko & Altmann 2010). Strong habits also have been formed as humans repeatedly navigate through virtual mazes (Marchette et al. 2011) and rats through actual mazes (Packard & Goodman 2013). Echoing insensitivity to reward, strong habits in these paradigms often emerge as errors, reflecting persistent responding despite task changes in the correct, rewarding outcome. Habits also emerge as chunked responses into a unit (Dezfouli & Balleine 2013, Graybiel 1998), which is related to the performance gains (reduced response time, increased accuracy) in many sequential learning tasks (e.g., Lungu et al. 2014). In addition, eye-tracking measures have been used to capture visual attention to task structures triggering habitual responses (Patsenko & Altmann 2010).

The strength of everyday habits is typically assessed from people's self-reports. One method is to combine self-reported ratings of behavioral frequency with ratings of the stability of the performance context, reflecting the logic that habits represent the repeated pairing of responses and recurring context cues (Galla & Duckworth 2015, Wood & Neal 2009). Using an alternative approach, the Self-Report Habit Index estimates habit strength from a questionnaire measure of the experience of automaticity and frequency of past performance (Verplanken & Orbell 2003), which has further been streamlined to measure only automaticity (Gardner 2014, Gardner et al. 2012). However, as Labrecque & Wood (2015) noted, experienced automaticity measures often fail to assess triggering contexts and may capture automation more generally, as opposed to habit automaticity that guides performance with limited input from current intentions. However, these measures can successfully capture habit strength when triggering cues are present (P. Lin, W. Wood, and J. Monterosso, unpublished data). Perhaps the most valid assessments of everyday habit strength involve reaction time measures of the accessibility of the habitual response given exposure to associated context cues (e.g., Neal et al. 2012).

In summary, assessments of habit formation rest most importantly on evidence that responses are insensitive to changes in rewarding outcomes. Habit formation has been documented in a variety of laboratory tasks using a variety of behavioral assays. For everyday habits, self-report assessments can capture habit strength, although direct assessment of context-response associations is probably most effective.

COMPUTATIONAL MODELS

Computational models offer detailed accounts of the cognitive processes that support habit learning and performance. We selectively focus on models that incorporate habit-like control systems as well as deliberate goal pursuit. Instead of making the simplified assumption that a behavior is either goal directed or habitual, these models explore how adaptive behavior can emerge from the interplay of different models of action control. Competing ideas have been proposed about how the different action controllers work together to produce a response.

In Cooper et al.'s (2014) goal circuit (GC) model, goals structure the learning of habits and control their expression. The GC model is an artificial neural network composed of two interlinked

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

habit and goal subnetworks. The habit system, which was originally proposed by Botvinick & Plaut (2004), selects actions in a bottom-up manner based on the current stimulus environment and on internal feedback about the network's previous state. This response selection process is biased by input from the goal system. A habit develops gradually as the network repeats the same sequence of responses while learning to attain goals in a particular environment. Eventually, the habit system becomes capable of performing a sequence autonomously without goal input. In addition to guiding learning in the habit system, the goal network enables top-down control over habitual action sequences, as when a person deliberately overrides a habitual response.

Taatgen et al. (2008) developed a model within the ACT-R (adaptive control of thought–rational) cognitive architecture that shows how behavioral control shifts from an internal, declarative task representation to environmental cues when acquiring a new action routine. Initially, explicit task knowledge is used to control behavior in a goal-directed manner. With practice, explicit memory retrieval is gradually transformed into a process by which perceptual cues trigger the relevant action directly. This proceduralization of explicit knowledge accounts for performance improvements during skill learning. Additional learning is possible by combining stimulus-cued productions of sequences into a single new production. The creation of new, specialized knowledge structures or chunks that can be used more efficiently is a central element in cognitive theories of skill acquisition (Newell 1990).

In cognitive neuroscience, the prevalent theoretical perspective is that goal-directed actions and habits can be described by different classes of RL (Daw et al. 2005, Dolan & Dayan 2013). Goal-directed, or model-based, learning is a computationally demanding process of mental simulation and planning. Using this approach, an agent computes on the fly which action maximizes long-run cumulative reward. Habit formation, in contrast, relies on model-free RL involving trial-and-error learning to estimate and store the long-run values of actions that are available in the different states or contexts. Actions are then chosen based on the stored or cached action values, reflecting predictions about future reward. Model-free control lacks the flexibility of model-based learning because short-term changes in the reward value of an action outcome have only limited effect on the cached value. Thus, model-free RL theoretically captures a key property of habits—insensitivity to changes in reward. Both types of learning are driven by prediction errors, with model-based learning capturing the discrepancy between the current state and the expected one and model-free learning capturing the difference between predicted and experienced reward.

In an initial dual-system RL model, Daw et al. (2005) assumed a competitive, winner-take-all mechanism in which the habit system or the goal system gained control over action, depending on which system provided the more reliable estimates of action values. However, subsequent investigations favor a more dynamic integration in which both systems contribute to the computation of action values according to their relative reliabilities (Lee et al. 2014). A related proposal involves Bayesian model averaging that takes into account prediction accuracy and model complexity. In this view, with experience, goal-directed actions are replaced by habits because the habit system becomes increasingly reliable and favored over the computationally more complex goal-directed system (FitzGerald et al. 2014). Furthermore, is it likely that model-based judgments are implemented only selectively, given the psychological costs of planning (see Habit and Deliberation section)? In recognition, a number of contingent RL systems engage model-based processing conditional on trade-offs between accuracy and efficiency (Keramati et al. 2011, Pezzulo et al. 2013).

Beyond integrating independently computed action values, some RL models assume a direct influence of goal values on habit learning. For example, in Gershman et al.'s (2014) two-step decision task, participants changed their choice preferences in the first task step after independently learning about second-step reward contingencies. To account for this finding, Gershman and colleagues argued that the model-based system simulated a complete two-step decision process, and

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

the model-free system learned from the simulations. Similarly, Pezzulo et al. (2013) proposed that model-based planning can update and improve the value estimates of the model-free system. By enabling goals to influence the action values represented in habit learning, these models promote the formation of habits compatible with goals.

The RL approach can account not only for the insensitivity of habits to changes in reward value but also for the chunking feature of habit automaticity. Theories of hierarchical RL show that it can be advantageous to concatenate individual actions and treat the sequence as a single response unit or chunk (Botvinick & Weinstein 2014) because the faster responding enabled by chunking can lead to greater average reward (Keramati et al. 2011). The chunked units could be deployed in a goal-directed (Dezfouli & Balleine 2012) or model-free manner on the basis of the reward history (Botvinick et al. 2009).

Questions remain, however, about the appropriateness of equating model-free learning with habit processes. Dezfouli & Balleine's (2012) proposal that habits are action chunks that are acquired and controlled through model-based processes marks a radical departure from the common RL assumption that habits are the result of model-free learning. Even more significant, individual-difference studies have reported that the strength of model-free learning was unrelated to habit formation and insensitivity to the value of the task outcome (Friedel et al. 2014, Gillan et al. 2015). Instead, outcome insensitivity was greater among participants evidencing little model-based learning. It is possible that the standard two-stage RL decision task does not capture the process that produces outcome-insensitive habits, perhaps because participants have to choose between multiple outcomes with varying reward rates (see discussion in the Associative and Reward Mechanisms in Habit Learning section). Thus, model-free learning in these tasks might reflect other stimulus-driven strategies such as simple decision heuristics instead of habits. Future research on model-free processes may need to develop new experimental tasks to better capture habits.

In summary, habits form as a product of repeated behaviors in the service of goal pursuit. Learning in the habit system may proceed independently or be guided by the goal system. Recent theories suggest that rather than being independent action controllers, habit and goal systems integrate in ways that reflect the reliability of each system and the costs of planning. In some contingent RL models, the habit system serves as an efficient default, and people plan only when motivated and able.

NEUROBIOLOGY OF HABITS

From a dual-system perspective, a fundamental objective is to identify brain regions whose activity is uniquely associated with habitual and goal-directed behavior, respectively. Current neuroscientific research is guided, to an increasing extent, by the computational RL models that we discussed in the previous section. Thanks to the rapid advancement of functional neuroimaging with human subjects, it is now possible to relate computational dual-process models to brain functioning at increasingly fine-grained levels of analysis. The emerging picture is one of a neurocognitive system that integrates the computations of partially overlapping neural systems of habitual and goal-directed control.

Neural Systems Associated with Habits

Research conducted with rodents, monkeys, and humans has provided converging evidence that habitual and goal-directed behaviors are mediated by neural circuits that link cortical brain areas and the basal ganglia (BG), a collection of subcortical nuclei. These circuits are organized as anatomically separate reentrant loops, two of which are essential for deliberate and habitual responding (Burton et al. 2015, Yin & Knowlton 2006). The associative cortico-BG loop supports

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

working memory functions and goal-directed actions and links the prefrontal cortex (PFC) with two striatal BG regions, the caudate nucleus and the anterior putamen. The sensorimotor loop underlies automatic, habitual behaviors and connects the somatosensory and motor cortex with the medial and posterior putamen. Though anatomically separate, the two loops can interact, for example, through spiraling dopaminergic connectivity (Haber et al. 2000).

Animal learning studies have demonstrated the importance of the sensorimotor loop for habitual responding. Rats do not acquire a lever-pressing habit when their dorsolateral striatum (DLS), the equivalent of the primate putamen, is lesioned prior to lever-press training for sucrose. Even after extended training, these rats continued to be goal directed and pressed the lever less frequently when sucrose was devalued (Yin et al. 2004). Furthermore, when the DLS was inactivated pharmacologically after a lever-pressing habit was acquired, outcome sensitivity was reinstated (Yin et al. 2006).

Suggesting that the DLS is involved in the chunking of individual actions into a sequence, electrical recordings from neurons in the DLS of rats exhibited a task-bracketing pattern of activity during habitual runs through a maze—high neuronal activity at the beginning and end of a run, with lower activity in-between. Task bracketing emerged when the learned behavior was still goal directed (Smith & Graybiel 2013), which indicates that habits develop in parallel with goal-directed learning but do not influence overt behavior early in training. By contrast, when goal-directed control is abolished by lesioning associated brain regions such as the rodent posterior dorsomedial striatum (DMS, corresponding to the caudate nucleus in primates) or the prelimbic medial PFC, behavior immediately comes under habitual control (Killcross & Coutureau 2003, Yin et al. 2005).

Neuroimaging research with human participants implicates similar networks of brain regions. Practicing sequences of finger movements for days or weeks decreased brain activation in areas associated with goal-directed control [e.g., premotor and prefrontal cortical areas, anterior cingulate cortex (ACC), and associative BG territories] and increased activation in the sensorimotor network, including the putamen (e.g., Lehéricy et al. 2005, Steele & Penhune 2010). Participants who developed a lever-pressing habit for potato chips and candy over three days of training showed similar increases in activity in the sensorimotor striatum (posterior putamen) both within practice days as well as across days (Tricomi et al. 2009). Neuroimaging studies of motor sequence learning further confirmed the role of the sensorimotor striatum in chunk formation, along with a frontoparietal network and the mediotemporal lobes (Lungu et al. 2014).

The sensorimotor loop is critical for habit learning, and an extended network of brain regions modulates its activity. Motivational influences on habit acquisition are mediated by ascending dopamine projections from the substantia nigra pars compacta to the dorsal striatum that modulate habit plasticity at corticostriatal synapses (Balleine & O'Doherty 2010). Lesioning this nigrostriatal pathway in rats disrupted habit formation (Faure et al. 2005). Habit acquisition also was impeded by lesions of the amygdala central nucleus, most likely due to its effect on substantia nigra pars compacta function (Lingawi & Balleine 2012). Finally, rodents' infralimbic cortex, a medial PFC region, directly participates in the formation of a habit and is required for its expression (Killcross & Coutureau 2003, Smith & Graybiel 2013).

Whether the sensorimotor loop is responsible for long-term habit storage remains unclear. After six months of practicing sequences of joystick movements, monkeys that had their sensorimotor loop disrupted pharmacologically were not impaired in the expression of sequence knowledge and still executed overlearned sequences faster and more accurately than random control sequences (Desmurget & Turner 2010). One explanation is that, with extensive practice, habit learning is consolidated in cortical brain areas (Atallah et al. 2007). This possibility fits Penhune & Steele's (2012) conclusion that long-term representations of learned skills are encoded in a network of

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

motor cortical regions. In their research, delayed recall of a motor sequence engaged not the BG but rather cortical regions (primary motor and premotor cortices and parietal lobe). This idea is supported further by the finding that participants who practiced motor sequences for more than six weeks showed, late in training, neural specialization in cortical motor areas [primary motor and premotor cortices and the supplementary motor area (SMA)] (Wymbs & Grafton 2014).

In summary, converging evidence implicates the sensorimotor cortico-BG loop as the core neural substrate of habit learning and performance. Whether the BG is also required for the long-term retention of habits is a matter of current debate.

Neural Systems that Integrate Habit and Goal-Directed Action Control

Based on computational RL theory, a stream of neuroimaging research has evaluated performance in sequential decision tasks that elicit both habit-based (model-free) learning and prospective planning about outcomes (model based). The results are broadly consistent with the mapping of habitual and goal-directed control onto the associative and sensorimotor cortico-BG loops. For example, Wunderlich et al. (2012) related activity in the posterior putamen to cached action values (habits) that were acquired through extensive training, and activity in the anterior caudate nucleus was related to values used in model-based planning. Similarly, Lee et al. (2014) reported that the posterior putamen, SMA, dorsomedial PFC, and dorsolateral PFC encoded model-free action values, whereas model-based values were associated with activity in orbitofrontal cortex (OFC) and medial PFC as well as the ACC (see also de Wit et al. 2009, Valentin et al. 2007).

Recent dual-system RL models propose that response selection is based on action values represented in ventromedial prefrontal cortex (vmPFC) that are jointly determined by a model-free (habit) and prospective planning controllers (e.g., Daw et al. 2011). This integration of model-based and model-free value signals is thought to be conducted by an arbitrator associated with activity in the inferior lateral PFC, frontopolar cortex, and ACC (Lee et al. 2014). How exactly such an arbitrator regulates the contribution of each system is still largely unknown. According to one analysis, shifts in response strategy are achieved primarily by strengthening or inhibiting the influence of the model-free habit system (Lee et al. 2014). Other findings suggest dynamic changes in both the habit system and the goal system (Gremel & Costa 2013).

Consistent with the idea that habits develop as people pursue goals, recent evidence suggests that multiple brain regions participate in both goal-directed and habitual control. For example, Lee et al. (2014) found with the multi-step decision task that two regions, the SMA and dmPFC, represented both model-free and model-based values. Similarly, Gremel & Costa (2013) trained rats to lever press for sucrose using either a habitual or a goal-directed strategy. They reported that a large proportion of neurons in DLS, DMS, and OFC participated in both habitual and goal-directed responding, and that the relative engagement of neurons in these areas depended on the current response strategy.

In summary, research guided by RL theory has identified the neural substrates of model-based (goal-directed) and model-free (habitual) control. These neural systems are partially overlapping, and their computations are integrated by brain regions that regulate the relative influence of the two modes of behavioral control.

FACTORS THAT SHIFT THE BALANCE BETWEEN HABITS AND DELIBERATE GOAL PURSUIT

Given the cognitive and neural features that differentiate habitual responding from more deliberate pursuit of goals, action control proceeds by balancing these and other processes (e.g., Pavlovian

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

HABIT SLIPS

Habit slips, or errors of inadvertent habit performance, occur primarily when an intended action and a habit share a performance context or some action component (Norman 1981). In daily life, people appear to slip up by performing unwanted habits about six times a week, especially when their attention is diverted from the task at hand (Reason 1979). For example, habit slips underlie errors in responding to innocent-appearing email phishing attacks (Vishwanath 2015). In laboratory tests, habits were most likely to be performed inadvertently when goal-directed control was taxed by, for example, advanced age or performing a secondary task (e.g., de Wit et al. 2014, Ruh et al. 2010).

Moving beyond the truism that people make errors when not attending to what they are doing, research has identified several sources of habit slips. Slips arising from failures in planning reflect limits in motivation or knowledge about completion of task goals, and they are most likely at decision points such as the end-of-task subroutines (Norman 1981, Reason 1979). However, habit slips also emerge from failures in automated, habitual guides to performance, as when degraded or forgotten representations of task context occur within a sequence of well-learned actions (Botvinick & Bylsma 2005). In still another analysis, habit slips arise from normal processing, especially open-loop action control in which habits are executed ballistically once they are launched, even when they are not the optimal response (Dezfouli et al. 2014, Orbell & Verplanken 2010). Thus, habit slips reflect failures to select the correct action through top-down control or bottom-up activation as well as ballistic habit performance.

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

conditioning; de Wit & Dickinson 2009). Tipping the balance between habits and goal pursuit are factors such as distraction (see Habit Slips sidebar), time pressure, limited task ability, or limited willpower (for a review, see Wood et al. 2014). These factors drive action control by pitting efficiency in processing against more effortful reliable processing (Evans & Stanovich 2013). That is, people act on strong habits when they lack the ability and motivation to engage the central executive in deliberation or, in RL terminology, when pressured by the time costs of model-based planning.

Considerable research reveals increased habit performance with impaired executive functioning. For example, when willpower was reduced by previously performing a taxing decision-making task, participants did not tailor their responses to their current circumstances but instead fell back on strongly habitual choices (Neal et al. 2013, Vohs et al. 2005). Also, when distracted by performing a demanding task, participants completing a categorization task increased the use of stimulus-response strategies over rule-based ones (Foerde et al. 2006), and participants completing a multistep decision task increased model-free responding (Otto et al. 2013a). Furthermore, in individual-difference paradigms, older adults and those with lower cognitive-control abilities were less able to leverage higher-order goal representations for model-based responding in order to overcome habitual, model-free solutions to a variety of tasks (de Wit et al. 2014, Otto et al. 2015). In like manner, participants possessing low spatial perspective-taking ability, after practicing navigating a maze, used more habitual and less goal-directed navigation strategies (Marchette et al. 2011).

Stress and drug addiction are of particular interest because of the multiple routes by which they tip the balance toward habits and away from deliberate decision making. As we explain in the next sections, these factors not only impede executive processes but also perhaps promote habit learning.

Stress

Acute as well as chronic stress can increase people's reliance on habits (Schwabe & Wolf 2013). For example, participants exposed to a combination of physical and psychosocial stressors (immersing a

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

hand into ice water while being monitored by a stranger and videotaped) after instrumental learning acted more habitually and were less sensitive to changes in the value of task rewards (Schwabe & Wolf 2010). This stress-induced shift toward habits is due in part to stress impeding deliberate action control. At the neural level, the shift toward habitual behavior was accompanied by decreased activity in OFC and medial PFC, brain regions associated with goal-directed learning (Schwabe et al. 2012). In sequential decision-making tasks, acute stress selectively attenuated deliberate, model-based control and promoted habit performance in vulnerable participants—those with low working-memory capacity (Otto et al. 2013b) or high levels of chronic stress (Radenbach et al. 2015). Similarly, in a study of visual classification learning, stressed participants were biased toward relying on a habit-linked procedural learning strategy at the expense of explicit learning (Schwabe & Wolf 2012). These results could reflect simply the breakdown of higher-order decision-making functions under stress, or stress could lead to a shift in the allocation of cognitive resources so that people fall back on habits and other strategies to prevent unreliable performance. From an RL perspective, stress shifts the balance toward habits over lengthy planning by increasing experienced time pressure (Doll et al. 2012).

Along with impeding deliberate thought, stress also might promote habit acquisition. Research with rodents suggests that stress can, under specific conditions, facilitate habit learning through mechanisms associated with dorsolateral striatal function (Dias-Ferreira et al. 2009). In humans, however, stress does not clearly affect habit formation. To isolate stress effects on habit learning, stress is induced before training, and learning is assessed after acute stress effects have dissipated so that these do not affect performance. Administering stress hormones led to improved learning in a simple stimulus-response task but had no effect on habit learning in a virtual radial maze (Guenzel et al. 2014a). Furthermore, a combination of pretraining physical and psychosocial stressors actually impaired habitual performance at maze navigation, albeit only in male participants (Guenzel et al. 2014b).

In summary, neurophysiological responses to stress increase habitual responding by impeding deliberate action control and, potentially, by promoting habit formation. More generally, stress research highlights the benefits of habits for rescuing performance. In support of this functional role, stressed participants' task performance was impaired to the extent that they attempted to engage goal-directed neural systems (Schwabe & Wolf 2013). Given the ready acquisition and performance of habits, they provide a useful default when threat and pressure derail more thoughtful responding.

Addiction

From a habit perspective, the path to drug addiction involves not a pathological motivation for drugs but rather a shift from goal-directed to habitual drug seeking and consumption (Everitt 2014, Hogarth et al. 2013). Initial drug seeking is voluntary and reflects the hedonic value of the drug. Through instrumental learning with drug rewards, context cues rapidly become associated with drug use. In addition, Pavlovian mechanisms contribute to the want for drugs and for cue-evoked cravings (Berridge 2007). These various learning mechanisms are involved in the cues that come to trigger drug seeking and consumption independently of the drug outcome, much as people repeat habits with limited sensitivity to goals and outcomes (Zapata et al. 2010). Phenomenologically, the addict no longer likes the drug yet uses it compulsively, often despite intentions to quit. Drug-outcome insensitivity is promoted further as repeated exposures build tolerance to rewarding drug effects.

Drug use promotes habit formation in part by impairing goal-directed control. In illustration, study participants who had consumed alcohol responded habitually and continued to choose

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

chocolate in a repeated food-choice task despite having just eaten three chocolate bars (Hogarth et al. 2012a). Comparable findings have emerged with chronic addicts abstinent at test. For example, participants who were obese, obsessive-compulsive, or dependent on methamphetamine (abstinent at test) showed compromised goal-directed learning at a decision-making task, along with a maladaptive reliance on habits (Voon et al. 2015). Furthermore, these responses in chronic addicts were associated with neural markers of lower gray matter volumes in the caudate, medial OFC, and lateral PFC. Also, alcohol-dependent participants responded habitually after rewards had been devalued in an associative learning task, and greater reliance on habits was associated with reduced engagement of brain areas implicated in goal-directed action (vmPFC) and increased engagement of areas implicated in habit learning (posterior putamen; Sjoerds et al. 2013). Even when not under drug influence, simply being in the presence of cues repeatedly associated with drug exposure can disrupt goal-directed responding (Ostlund et al. 2010). In general, goal-directed impairments increasingly narrow addicts' behavioral repertoires onto drug habits by restricting their capacity for intentionally selecting alternative actions.

Drug use also promotes habit responding through neurobiological processes that sensitize users to the incentive properties of drugs. Drug rewards appear to engage habits more rapidly than other reinforcers (see review in Everitt 2014). Stimulants in particular accelerate and consolidate the development of drug use habits, speeding the neural shifts from associative to sensorimotor areas typically found with habit formation. The accelerated formation of habits hastens the transition from initial or occasional user to addict.

In summary, drug exposure hijacks the habit learning system by exerting a continuous pressure in favor of habitual, context-driven behavior and away from the evaluation of the outcomes of action. As a result of these habitual and deliberative processes, drug use escalates so that people ultimately seek drugs compulsively (Redish et al. 2008).

INFERENCES ABOUT THE CAUSES OF HABIT PERFORMANCE

People often are aware of their habitual responses although they are largely unaware of the cuing mechanism that activates habits. Given this limited introspective access, people's explanations for their habitual responses are largely post hoc accounts. According to classic social psychology theories, when internal cues to action are weak, ambiguous, or uninterpretable, people infer what their motivations must be from observing their behavior and the context in which it occurs (Bem 1972).

The simple frequency of habit performance plausibly implies strong, consistent underlying motives. Such inferences could be correct in an historical sense, because people might accurately remember the goals that initially guided habit formation. However, they will not be correct concerning current habit performance. In support of such an inference, participants with stronger habits were more certain about their behavioral intentions and perceived the behavior as guided more by their goals than were participants with weaker habits, when in fact the opposite was true—intentions and goals were particularly poor predictors of strongly habitual behaviors (Ji & Wood 2007, Neal et al. 2012). People infer that goals underlie a range of habitual behaviors, even habits of compulsive drug seeking and use. Everitt & Robbins (2005) suggested that addicts' experience of wanting a drug is not a precursor of consumption but rather a post hoc rationalization of compulsive drug-use habits. Similarly, obsessive-compulsive disorders may originate in excessive habit formation, with irrational threat beliefs then inferred to explain the compulsively repeated behaviors (Gillan & Robbins 2014). Although people infer intentionality for a wide variety of automatically cued responses, when internal cues are strong and unambiguous, causal inferences

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

are unnecessary. Then, inconsistencies between habits and goals might just be labeled as such (e.g., "I can't help it, it's a habit").

The inference of motives behind habit performance is represented in some computational models as non-goal-mediated routine responding giving rise to goal representations. For example, in Sun et al.'s (2001) CLARION (Connectionist Learning with Adaptive Rule Induction Online) model, habitual responses that are controlled through bottom-up procedural knowledge can, over time, come to be represented in top-down rules via a rule extraction-refinement algorithm. In Cooper et al.'s (2014) goal circuit model, stimulus-driven habitual responses can activate goal representations that subsequently provide input to habits. In this way, active goals may be a consequence, rather than a precursor, of habitual action. From the perspective of RL models, this inference might occur through the model-based system adopting the values expressed in habit learning, perhaps to reduce computational overhead (Doll et al. 2012).

Goal inferences are sparked not only by the simple frequency of habit performance but also by the positive affect associated with many habits. Habits are likely to be favored due to the ease with which they can be performed compared with alternatives. Thus, consumers value using existing products and services over new ones because of the difficulty of learning new usage behaviors (e.g., computer program updates; Murray & Häubl 2007). Habits also are likely to be viewed positively due to the fluency or speed and ease of processing associated with frequently performed behaviors. High fluency is experienced as positive in part because it signals familiarity over uncertainty and success at processing and understanding, and this positive affect generalizes to evaluation of the activity (Reber et al. 2004). Habit inferences thus exploit a psychological calculus that favors what feels easy because it is well practiced over what feels more difficult because it is new. Being favorably disposed toward habits, people may infer that they intended to perform the response.

Although the inferences that follow habit performance may not be accurate descriptions of the mechanisms actually generating action, such inferences sometimes contribute to well-being. Repeated behaviors, such as students' choices of the same seat in a classroom, heighten feelings of comfort, confidence, and control despite that these choices might initially be random (Avni-Babad 2011). Furthermore, habit performance may promote coherence or comprehensibility of experiences and thus enhance meaning in life (S. Heintzelman and L. King, unpublished observation). However, inferences about habits are not always beneficial, and the transparency of habit knowledge to introspection can lead people to underestimate its usefulness. For example, when highly motivated to perform well, participants with good procedural knowledge at a task overrode their habits and responded thoughtfully, despite that this impaired task performance (L. Carden, W. Wood, D. Neal, and A. Pascoe, unpublished observation).

In summary, people may explain habit performance, even addictive habits, by inferring relevant goals and intentions. Despite being largely erroneous, the inference that habits were intended may seem intuitively plausible given response frequency. Also relevant, switching costs can discourage deviating from habits, and experienced fluency can increase liking for them.

CHANGING HABITS

Unwanted habits are at the root of many failed attempts at behavior change. Evidence comes from Webb & Sheeran's (2006) meta-analysis of 47 studies that successfully used persuasive appeals and other interventions to change participants' intentions. The changed intentions, however, only yielded change in behaviors that participants performed sporadically (e.g., course enrollment) and not in behaviors that could be repeated into habits (e.g., seat belt use). Even the largely effective implementation intentions, or if-then plans to act on intentions at particular times and places (Gollwitzer & Sheeran 2006), are not successful at controlling many strong habits (Webb et al. 2009).

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

From a habit perspective, difficulties in changing established behavior patterns do not reflect people's continuing desire to perform the old behavior or a failure of willpower. The central challenge is that old habits continue to be activated automatically by recurring environmental cues (J. Labrecque and W. Wood, unpublished observation; Walker et al. 2014). Even after new habits have been learned, old memory traces are not necessarily replaced (Bouton et al. 2011). Relapse can occur when old habit memories are activated by prior routines and other context cues.

Addressing the role of habit learning is a central challenge for the next generation of behavior change interventions (Marteau et al. 2012, Rothman et al. 2015). In response to this challenge, interventions can be designed to (*a*) impede the automated cueing of old, unwanted habits as well as (*b*) promote the repetition of a new, desired behavior into a habit.

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

Interventions to Impede Unwanted Habit Performance

To reduce interference from old habits, behavior change interventions can incorporate mechanisms of inhibition. In research investigating the spontaneous inhibition of unwanted habits in daily life, the most successful strategy involved thinking, "Don't do it," and being mindful of slip-ups (Quinn et al. 2010). This strategy also successfully controlled habit errors when participants were instructed to use it in an experimental task. As expected, it worked by enhancing cognitive control, not by decreasing habit strength (Quinn et al. 2010). Also effective is tying inhibitory plans to the cues that activate unwanted habits (e.g., "After dinner, I'll skip dessert as usual and substitute fruit"; Adriaanse et al. 2010).

Interference from old habits also can be reduced by changing cues in performance environments. Animal research suggests that habit performance is especially impaired when contexts shift, with goal-directed responding transferring more successfully across contexts (Thrailkill & Bouton 2015). One way to change habit cues is through managing exposure. For example, unhealthy eating habits can be curbed by increasing the salience or accessibility of healthy foods (Sobal & Wansink 2007). A study illustrated how people do this in daily life: At all-you-can-eat Chinese buffets, patrons with lower body mass index used chopsticks, chose small plates, put napkins on their laps, and sat with their sides or backs to the buffet (Wansink & Payne 2012). Another way that habit cues change is through naturally occurring life transitions, such as when people switch jobs or move house. Habit discontinuity interventions capitalize on this reduced exposure to cues that trigger old habits (Thøgersen 2012, Verplanken et al. 2008, Walker et al. 2014). Life transitions that alter habit cues can provide a window of opportunity to act on new intentions without competition from old habits (Wood et al. 2005).

Interventions to Promote Formation of Desired Habits

By encouraging the formation of new habits, behavior change interventions can be designed to habitize a new behavior so that it is maintained despite short-term desires and temptations. In an experiment illustrating how healthy habits can arm people against succumbing to food temptations (P.-Y. Lin, W. Wood, and J. Monterosso, unpublished data), participants learned in a computerized task to avoid chocolate (study 1) or to choose carrots (study 2). When later given options of eating unhealthy treats, participants continued to make habitual choices, at least when context cues automatically triggered the healthy habit. Also, echoing the factors that balance habits and goals, participants were most likely to act on their healthy habits when they lacked the willpower to deliberate about food choices.

Despite the promise of habit formation in maintaining desired behaviors, only a few interventions with nonclinical populations have built on the three pillars of habit formation–repetition in

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

stable contexts with appropriate reward schedules (see Lally & Gardner 2013, Rünger & Wood 2015). The importance of frequent responding was illustrated in Lally et al.'s (2010) field experiment in which repetitions of a simple health behavior (e.g., walking after dinner) required from 18 days to as many as 254 days in the same context to become habitual and performed without thinking. For exercise, Kaushal & Rhodes (2015) estimated that going to the gym became automatic within six weeks, assuming visits of four times per week. Unfortunately, few health or other behavioral interventions have adapted interval reward schedules to facilitate habit formation (Burns et al. 2012). However, the importance of stable cues was demonstrated in interventions using tooth brushing to cue dental flossing and form flossing habits (e.g., Judah et al. 2013, Orbell & Verplanken 2010). The few interventions built on the three components of habit formation have yielded promising results for weight loss (Carels et al. 2014, Lally et al. 2008) and consumption of healthy food in families (Gardner et al. 2014).

Interestingly, in a habit formation intervention, electronic reminders to perform a desired behavior increased repetition but also impeded automaticity and learning of context-response associations (Stawarz et al. 2015). Perhaps reminders engaged deliberate decision making that impaired learning context–response associations (see Associative and Reward Mechanisms in Habit Learning section). More passive reminders in the form of physical signs, although helpful in prompting initial repetition and habit formation, ultimately lost potency over time (Tobias 2009).

Additional evidence of the utility of habit formation comes from interventions that did not specifically target habits. For example, forming gym-going habits enabled new members of a health club to sustain working out (Armitage 2005), and forming nonsmoking habits enabled former smokers to remain abstinent a year after the end of a smoking cessation program (Baldwin et al. 2006). In a study of regular exercisers, approximately 90% had a location or time cue to exercise, and exercising was more automatic for those who exercised in a routine way and were cued by a particular location (Tappe et al. 2013).

Research on self-control also suggests the usefulness of habits for maintaining desired behaviors. People with high trait self-control do not appear to attain goals through inhibition of problematic desires but instead through forming habits that allow them to achieve goals without experiencing unwanted temptations (Galla & Duckworth 2015, Hofmann et al. 2012). Trait self-control generally fosters proficiency at performing tasks that require automation (de Ridder et al. 2012). Nonetheless, everyone tends to fall back on performing habits—both good and bad—when they lack the capacity or motivation to make decisions to act in nonhabitual ways (Neal et al. 2013).

In summary, through combating unwanted habits and ensuring that desired behaviors are repeated in ways that promote habit formation, interventions can promote adoption of behaviors that endure over time. These interventions adapt the habit strategies that people with effective self-control use in their daily lives to ensure successful goal attainment.

CONCLUSION

The current state of the science on habits has provided the definition that James (1890) requested, overturned behaviorists' conceptions of simple stimulus-response associations, and placed habits within broader models of goal-directed action. Habits reflect associative learning and the formation of context-response associations in procedural memory. Once habits form, perception of the context automatically brings the response to mind, and people often carry out that response. As habits strengthen, they gradually become independent of the incentive value of their consequences, and neural activation shifts from associative toward sensorimotor cortico-striatal brain regions. When repeated in a sequence, habitual responses also may be chunked together and activated as a unit. In short, our review provides a framework for understanding, predicting, and changing that

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

common component of everyday life in which behavioral control has been outsourced onto the context cues contiguous with past performance.

Although habits are largely insensitive to changes in goal structure and value, they interact in three different ways with deliberate goal pursuit. First, habits form in daily life as people pursue goals by repeating actions in particular performance contexts. Initially, goals and declarative task knowledge structure behavior. With repetition, responses and associated context cues are captured in procedural learning systems. Goals also may contribute to habit formation through heightening attention to certain stimuli and identifying the value of action outcomes. Given the profusion of direct and indirect connections between neural circuits underlying goal-directed and habitual (model-free) behaviors, goals can have a biasing influence on habit formation (Doll et al. 2012). Cross talk between habit systems and more deliberative action control, especially during habit formation, is consistent with an evolutionary history in which neural systems supporting more sophisticated planning capacities evolved on top of neural mechanisms associated with habits.

A second interface between goals and habits emerges after habits form. That is, habits provide an efficient baseline response that likely integrates with more effortful goal pursuit only when necessary, as when habits prove unreliable in a given context or when people are especially motivated and able to tailor responses to particular circumstances. Various factors impede people's ability to deliberate and thus tip the balance toward relying on habits, including time pressure, distraction, stress, and addiction. Addictive substances may in addition promote habit responding by accelerating habit learning.

A third way in which goals integrate with habits is through the explanations that people generate for their habits. Because habit automaticity is inaccessible to subjective experience, people must infer the reasons for such responses. A plausible inference for repeated behaviors is strong, consistent underlying motivations and goals.

The research we reviewed highlights a number of advantages to acting habitually. For example, habit knowledge is protected from short-term whims and occasional happenings, given that habits form through incremental experience and do not shift readily with changes in people's goals and plans. Also, by outsourcing action control to environmental cues, people have a ready response when distraction, time pressure, lowered willpower, and stress reduce the capacity to deliberate about action and tailor responses to current environments. Furthermore, habit systems are smart in the sense that they enable people to efficiently capitalize on environmental regularities.

As we noted at the end of the present review, understanding habits is important from the applied perspective of human health and welfare. Drug addictions and other compulsions appear to co-opt habit processes and reduce people's capacity to purposively guide their behavior. Lifestyle habits of poor diet and limited exercise are major contributors to chronic diseases. By building on an understanding of habit mechanisms, addiction treatments as well as interventions to change lifestyle behaviors may successfully disrupt these unwanted habits and help people to form more effective habits that meet their goals for healthy, productive lives.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors thank Peter Dayan, Sanne de Wit, Benjamin Gardner, Barbara Knowlton, David T. Neal, Yael Niv, Sheina Orbell, A. Ross Otto, Carol Seger, Kyle Smith, and Bas Verplanken

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

for their thoughtful comments on an earlier version of the article. This article was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed are those of the authors and do not necessarily reflect the Foundation's views.

LITERATURE CITED

Aarts H, Verplanken B, Van Knippenberg A. 1997. Habit and information use in travel mode choices. *Acta Psychol.* 96:1–14

Adriaanse MA, Oettingen G, Gollwitzer PM, Hennes EP, de Ridder DTD, de Wit JBF. 2010. When planning is not enough: fighting unhealthy snacking habits by mental contrasting with implementation intentions (MCII). *Eur. J. Soc. Psychol.* 40:1277–93

Armitage CJ. 2005. Can the theory of planned behavior predict the maintenance of physical activity? *Health Psychol.* 24:235–45

Atallah HE, Lopez-Paniagua D, Rudy JW, O'Reilly RC. 2007. Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. *Nat. Neurosci.* 10:126–31

Avni-Babad D. 2011. Routine and feelings of safety, confidence, and well-being. *Br. J. Psychol.* 102:223–44

Baldwin AS, Rothman AJ, Hertel AW, Linde JA, Jeffery RW, et al. 2006. Specifying the determinants of the initiation and maintenance of behavior change: an examination of self-efficacy, satisfaction, and smoking cessation. *Health Psychol.* 25:626–34

Balleine BW, O'Doherty JP. 2010. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35:48–69

Bayley PJ, Frascino JC, Squire LR. 2005. Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature* 436:550–53

Bem DJ. 1972. Constructing cross-situational consistencies in behavior: some thoughts on Alker's critique of Mischel. *J. Personal.* 40:17–26

Berridge KC. 2007. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology* 191:391–431

Betsch T, Haberstroh S, Glöckner A, Haar T, Fiedler K. 2001. The effects of routine strength on adaptation and information search in recurrent decision making. *Organ. Behav. Hum. Decis. Process.* 84:23–53

Botvinick MM, Bylsma LM. 2005. Distraction and action slips in an everyday task: evidence for a dynamic representation of task context. *Psychol. Bull. Rev.* 12:1011–17

Botvinick MM, Niv Y, Barto AC. 2009. Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition* 113:262–80

Botvinick MM, Plaut DC. 2004. Doing without schema hierarchies: a recurrent connectionist approach to normal and impaired routine sequential action. *Psychol. Rev.* 111:395–429

Botvinick MM, Weinstein A. 2014. Model-based hierarchical reinforcement learning and human action control. *Philos. Trans. R. Soc. B* 369:20130480

Bouton ME, Todd TP, Vurbic D, Winterbauer NE. 2011. Renewal after the extinction of free operant behavior. *Learn. Behav.* 39:57–67

Burns RJ, Donovan AS, Ackermann RT, Finch EA, Rothman AJ, Jeffery RW. 2012. A theoretically grounded systematic review of material incentives for weight loss: implications for interventions. *Ann. Behav. Med.* 44:375–88

Burton AC, Nakamura K, Roesch MR. 2015. From ventral-medial to dorsal-lateral striatum: neural correlates of reward-guided decision-making. *Neurobiol. Learn. Mem.* 117:51–59

Carels RA, Burmeister JM, Koball AM, Oehlhof MW, Hinman N, et al. 2014. A randomized trial comparing two approaches to weight loss: differences in weight loss maintenance. *J. Health Psychol.* 19:296–311

Colwill RM, Rescorla RA. 1985. Postconditioning devaluation of a reinforcer affects instrumental responding. *J. Exp. Psychol.: Anim. Behav. Process.* 11:120–32

Cooper RP, Ruh N, Mareschal D. 2014. The goal circuit model: a hierarchical multi-route model of the acquisition and control of routine sequential action in humans. *Cogn. Sci.* 38:244–74

Danner UN, Vries NK, Aarts H. 2007. Habit formation and multiple means to goal attainment: Repeated retrieval of target means causes inhibited access to competitors. *Personal. Soc. Psychol. Bull.* 33:1367–79

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. 2011. Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69:1204–15

Daw ND, Niv Y, Dayan P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8:1704–11

de Ridder DTD, Lensvelt-Mulders G, Finkenauer C, Stok FM, Baumeister RF. 2012. Taking stock of self-control: a meta-analysis of how trait self-control relates to a wide range of behaviors. *Personal. Soc. Psychol. Rev.* 16:76–99

de Wit S, Corlett PR, Fletcher PC, Dickinson A, Aitken MR. 2009. Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *J. Neurosci.* 29:11330–38

de Wit S, Dickinson A. 2009. Associative theories of goal-directed behaviour: a case for animal-human translational models. *Psychol. Res.* 73:463–76

de Wit S, van de Vijver I, Ridderinkhof KR. 2014. Impaired acquisition of goal-directed action in healthy aging. *Cogn. Affect. Behav. Neurosci.* 14:647–58

Desmurget M, Turner RS. 2010. Motor sequences and the basal ganglia: kinematics, not habits. *J. Neurosci.* 30:7685–90

Dezfouli A, Balleine BW. 2012. Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* 35:1036–51

Dezfouli A, Balleine BW. 2013. Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized. *PLOS Comp. Biol.* 9:e1003364

Dezfouli A, Lingawi NW, Balleine BW. 2014. Habits as action sequences: hierarchical action control and changes in outcome value. *Philos. Trans. R. Soc. B* 369:20130482

Dias-Ferreira E, Sousa JC, Melo I, Morgado P, Cerqueira JJ. 2009. Chronic stress causes frontostriatal reorganization and affects decision-making. *Science* 325:621–25

Dickinson A. 1985. Actions and habits: the development of behavioural autonomy. *Philos. Trans. R. Soc. B* 308:67–78

Dolan RJ, Dayan P. 2013. Goals and habits in the brain. *Neuron* 80:312–25

Doll BB, Simon DA, Daw ND. 2012. The ubiquity of model-based reinforcement learning. *Curr. Opin. Neurobiol.* 22:1075–81

Evans J, Stanovich KE. 2013. Dual-process theories of higher cognition advancing the debate. *Perspect. Psychol. Sci.* 8:223–41

Everitt BJ. 2014. Neural and psychological mechanisms underlying compulsive drug seeking habits and drug memories—indications for novel treatments of addiction. *Eur. J. Neurosci.* 40:2163–82

Everitt BJ, Robbins TW. 2005. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* 8:1481–89

Faure A, Haberland U, Condé F, El Massioui N. 2005. Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *J. Neurosci.* 25:2771–80

FitzGerald THB, Dolan RJ, Friston KJ. 2014. Model averaging, optimal inference, and habit formation. *Front. Hum. Neurosci.* 8:457

Foerde K, Knowlton BJ, Poldrack RA. 2006. Modulation of competing memory systems by distraction. *PNAS* 103:11778–83

Friedel E, Koch SP, Wendt J, Heinz A, Deserno L, Schlagenhauf F. 2014. Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Front. Hum. Neurosci.* 8:587

Galla BM, Duckworth AL. 2015. More than resisting temptation: Beneficial habits mediate the relationship between self-control and positive life outcomes. *J. Personal. Soc. Psychol.* 109:508–25

Gardner B. 2014. A review and analysis of the use of "habit" in understanding, predicting and influencing health-related behaviour. *Health Psychol. Rev.* doi: 10.1080/17437199.2013.876238

Gardner B, Abraham C, Lally P, de Bruijn GJ. 2012. Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *Int. J. Behav. Nutr. Phys. Act.* 9:102

Gardner B, de Bruijn GJ, Lally P. 2011. A systematic review and meta-analysis of applications of the Self-Report Habit Index to nutrition and physical activity behaviours. *Ann. Behav. Med.* 42:174–87

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

Gardner B, Sheals K, Wardle J, McGowan L. 2014. Putting habit into practice, and practice into habit: a process evaluation and exploration of the acceptability of a habit-based dietary behaviour change intervention. *Int. J. Behav. Nutr. Phys. Act.* 11:135

Gershman SJ, Markman AB, Otto AR. 2014. Retrospective revaluation in sequential decision making: a tale of two systems. *J. Exp. Psychol.: Gen.* 143:182–94

Gillan CM, Otto AR, Phelps EA, Daw ND. 2015. Model-based learning protects against forming habits. *Cogn. Affect. Behav. Neurosci.* 15:523–36

Gillan CM, Robbins TW. 2014. Goal-directed learning and obsessive-compulsive disorder. *Philos. Trans. R. Soc. B* 369:20130475

Gollwitzer PM, Brandstätter V. 1997. Implementation intentions and effective goal pursuit. *J. Personal. Soc. Psychol.* 73:186–99

Gollwitzer PM, Sheeran P. 2006. Implementation intentions and goal achievement: a meta-analysis of effects and processes. In *Advances in Experimental Social Psychology*, Vol. 38, ed. MP Zanna, pp. 69–119. San Diego, CA: Elsevier

Graybiel AM. 1998. The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* 70:119–36

Gremel CM, Costa RM. 2013. Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nat. Commun.* 4:2264

Guenzel FM, Wolf OT, Schwabe L. 2014a. Glucocorticoids boost stimulus-response memory formation in humans. *Psychoneuroendocrinology* 45:21–30

Guenzel FM, Wolf OT, Schwabe L. 2014b. Sex differences in stress effects on response and spatial memory formation. *Neurobiol. Learn. Mem.* 109:46–55

Haber SN, Fudge JL, McFarland NR. 2000. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J. Neurosci.* 20:2369–82

Hay JF, Jacoby LL. 1996. Separating habit and recollection: memory slips, process dissociations, and probability matching. *J. Exp. Psychol.: Learn. Mem. Cogn.* 22:1323–35

Hofmann W, Baumeister RF, Förster G, Vohs KD. 2012. Everyday temptations: an experience sampling study of desire, conflict, and self-control. *J. Personal. Soc. Psychol.* 102:1318–35

Hogarth L, Attwood AS, Bate HA, Munafò MR. 2012a. Acute alcohol impairs human goal-directed action. *Biol. Psychol.* 90:154–60

Hogarth L, Balleine BW, Corbit LH, Killcross S. 2013. Associative learning mechanisms underpinning the transition from recreational drug use to addiction. *Ann. NY Acad. Sci.* 1282:12–24

Hogarth L, Chase HW, Baess K. 2012b. Impaired goal-directed behavioural control in human impulsivity. *Q. J. Exp. Psychol.* 65:305–16

Holland PC. 2004. Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *J. Exp. Psychol.: Anim. Behav. Proc.* 30:104–17

Hommel B. 2009. Action control according to TEC (theory of event coding). *Psychol. Res.* 73:512–26

Hull CL. 1943. *Principles of Behavior: An Introduction to Behavior Theory*. New York: Appleton-Century

James W. 1890. *The Principles of Psychology*. New York: H. Holt

Ji MF, Wood W. 2007. Purchase and consumption habits: not necessarily what you intend. *J. Consum. Psychol.* 17:261–76

Judah G, Gardner B, Aunger R. 2013. Forming a flossing habit: an exploratory study of the psychological determinants of habit formation. *Br. J. Health Psychol.* 18:338–53

Kaushal N, Rhodes RE. 2015. Exercise habit formation in new gym members: a longitudinal study. *J. Behav. Med.* 38:652–63

Keramati M, Dezfouli A, Piray P. 2011. Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLOS Comp. Biol.* 7:e1002055

Khare A, Inman JJ. 2006. Habitual behavior in American eating patterns: the role of meal occasions. *J. Consum. Res.* 32:567–75

Killcross S, Coutureau E. 2003. Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex* 13:400–8

Labrecque JS, Wood W. 2015. What measures of habit strength to use? Comment on Gardner 2014. *Health Psychol. Rev.* doi: 0.1080/17437199.2014.992030

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

Lally P, Chipperfield A, Wardle J. 2008. Healthy habits: efficacy of simple advice on weight control based on a habit-formation model. *Int. J. Obes.* 32:700–7

Lally P, Gardner B. 2013. Promoting habit formation. *Health Psychol. Rev.* 7:S137–58

Lally P, Van Jaarsveld CHM, Potts HWW, Wardle J. 2010. How are habits formed: modelling habit formation in the real world. *Eur. J. Neurosci.* 40:998–1009

Lee SW, Shimojo S, O'Doherty JP. 2014. Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81:687–99

Lehéricy S, Benali H, Van de Moortele PF, Pélégrini-Issac M, Waechter T, et al. 2005. Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *PNAS* 102:12566–71

Liljeholm M, Dunne S, O'Doherty JP. 2015. Differentiating neural systems mediating the acquisition versus expression of goal-directed and habitual behavioral control. *Eur. J. Neurosci.* 41:1358–71

Lingawi NW, Balleine BW. 2012. Amygdala central nucleus interacts with dorsolateral striatum to regulate the acquisition of habits. *J. Neurosci.* 32:1073–81

Liu JL, Han B, Cohen DA. 2015. Associations between eating occasions and places of consumption among adults. *Appetite* 87:199–204

Lungu OL, Monchi O, Albouy G, Jubault T, Ballarin E, et al. 2014. Striatal and hippocampal involvement in motor sequence chunking depends on the learning strategy. *PLOS ONE* 9:e103885

Maher JP, Conroy DE. 2015. Habit strength moderates the effects of daily action planning prompts on physical activity but not sedentary behavior. *J. Sport Exerc. Psychol.* 37:97–107

Marchette SA, Bakker A, Shelton AL. 2011. Cognitive mappers to creatures of habit: differential engagement of place and response learning mechanisms predicts human navigational behavior. *J. Neurosci.* 31:15264–68

Marteau TM, Hollands GJ, Fletcher PC. 2012. Changing human behavior to prevent disease: the importance of targeting automatic processes. *Science* 337:1492–95

Miller GA, Galanter E, Pribram KH. 1960. *Plans and the Structure of Behavior*. New York: Holt, Rinehart, & Winston

Murray KB, Häubl G. 2007. Explaining cognitive lock-in: the role of skill-based habits of use in consumer choice. *J. Consum. Res.* 34:77–88

Neal DT, Wood W, Drolet A. 2013. How do people adhere to goals when willpower is low? The profits (and pitfalls) of strong habits. *J. Personal. Soc. Psychol.* 104:959–75

Neal DT, Wood W, Labrecque JS, Lally P. 2012. How do habits guide behavior? Perceived and actual triggers of habits in daily life. *J. Exp. Soc. Psychol.* 48:492–98

Neal DT, Wood W, Wu M, Kurlander D. 2011. The pull of the past: When do habits persist despite conflict with motives? *Personal. Soc. Psychol. Bull.* 37:1428–37

Newell A. 1990. *Unified Theories of Cognition*. Cambridge, MA: Harvard Univ. Press

Norman DA. 1981. Categorization of action slips. *Psychol. Rev.* 88:1–15

Orbell S, Verplanken B. 2010. The automatic component of habit in health behavior: habit as cue-contingent automaticity. *Health Psychol.* 29:374–83

Ostlund SB, Maidment NT, Balleine BW. 2010. Alcohol-paired contextual cues produce an immediate and selective loss of goal-directed action in rats. *Front. Integr. Neurosci.* 4:19

Otto AR, Gershman SJ, Markman AB, Daw ND. 2013a. The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol. Sci.* 24:751–61

Otto AR, Raio CM, Chiang A. 2013b. Working-memory capacity protects model-based learning from stress. *PNAS* 52:20941–46

Otto AR, Skatova A, Madlon-Kay S, Daw ND. 2015. Cognitive control predicts use of model-based reinforcement learning. *J. Cogn. Neurosci.* 27:319–33

Ouellette JA, Wood W. 1998. Habit and intention in everyday life: the multiple processes by which past behavior predicts future behavior. *Psychol. Bull.* 124:54–74

Packard MG, Goodman J. 2013. Factors that influence the relative use of multiple memory systems. *Hippocampus* 23:1044–52

Patsenko EG, Altmann EM. 2010. How planful is routine behavior? A selective-attention model of performance in the Tower of Hanoi. *J. Exp. Psychol.: Gen.* 139:95–116

Penhune VB, Steele CJ. 2012. Parallel contributions of cerebellar, striatal and M1 mechanisms to motor sequence learning. *Behav. Brain Res.* 226:579–91

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

Pezzulo G, Rigoli F, Chersi F. 2013. The mixed instrumental controller: using value of information to combine habitual choice and mental simulation. *Front. Psychol.* 4:92

Quinn JM, Pascoe A, Wood W, Neal DT. 2010. Can't control yourself? Monitor those bad habits. *Personal. Soc. Psychol. Bull.* 36:499–511

Radenbach C, Reiter AMF, Engert V, Sjoerds Z, Villringer A, et al. 2015. The interaction of acute and chronic stress impairs model-based behavioral control. *Psychoneuroendocrinology* 53:268–80

Reason JT. 1979. Actions not as planned: the price of automatization. In *Aspects of Consciousness*, ed. G Underwood, R Stevens, pp. 67–89. London: Academic

Rebar AL, Elavsky S, Maher JP, Doerksen SE, Conroy DE. 2014. Habits predict physical activity on days when intentions are weak. *J. Sport Exerc. Psychol.* 36:157–65

Reber R, Schwarz N, Winkielman P. 2004. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personal. Soc. Psychol. Rev.* 8:364–82

Redish AD, Jensen S, Johnson A. 2008. A unified framework for addiction: vulnerabilities in the decision process. *Behav. Brain Sci.* 31:415–37

Rothman AJ, Gollwitzer PM, Grant AM, Neal DT, Sheeran P, Wood W. 2015. Hale and hearty policies: how psychological science can create and maintain healthy habits. *Perspect. Psychol. Sci.* In press

Rubin G. 2015. *Better than Before*. New York: Random House

Ruh N, Cooper RP, Mareschal D. 2010. Action selection in complex routinized sequential behaviors. *J. Exp. Psychol.: Hum. Percept. Perform.* 36:955–75

Rünger D, Wood W. 2015. Maintenance of healthy behaviors: forming and changing habits. In *Behavioral Economics and Public Health*, ed. C Roberto, I Kawachi. Oxford, UK: Oxford Univ. Press. In press

Schwabe L, Tegenthoff M, Höffken O, Wolf OT. 2012. Simultaneous glucocorticoid and noradrenergic activity disrupts the neural basis of goal-directed action in the human brain. *J. Neurosci.* 32:10146–55

Schwabe L, Wolf OT. 2010. Socially evaluated cold pressor stress after instrumental learning favors habits over goal-directed action. *Psychoneuroendocrinology* 35:977–86

Schwabe L, Wolf OT. 2012. Stress modulates the engagement of multiple memory systems in classification learning. *J. Neurosci.* 32:11042–49

Schwabe L, Wolf OT. 2013. Stress and multiple memory systems: from "thinking" to "doing." *Trends Cogn. Sci.* 17:60–68

Seger CA, Spiering BJ. 2011. A critical review of habit learning and the basal ganglia. *Front. Syst. Neurosci.* 5:1–9

Sheeran P, Webb TL, Gollwitzer PM. 2005. The interplay between goal intentions and implementation intentions. *Personal. Soc. Psychol. Bull.* 31:87–98

Shiffrin RM, Schneider W. 1977. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychol. Rev.* 84:127–90

Sjoerds Z, de Wit S, van den Brink W, Robbins TW, Beekman ATF, et al. 2013. Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Transl. Psychiatry* 3:e337

Skinner BF. 1938. *The Behavior of Organisms*. New York: Appleton-Century-Crofts

Smith KS, Graybiel AM. 2013. A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron* 79:361–74

Sobal J, Wansink B. 2007. Kitchenscapes, tablescapes, platescapes, and foodscapes: influences of microscale built environments on food intake. *Environ. Behav.* 39:124–42

Squire LR, Zola-Morgan S. 1991. The medial temporal lobe memory system. *Science* 253:1380–86

Stawarz K, Cox AL, Blandford A. 2015. Beyond self-tracking and reminders: designing smartphone apps that support habit formation. In *Proc. 33rd annu. ACM Conf. Hum. Factors Comput.*, pp. 2653–62. New York: ACM

Steele CJ, Penhune VB. 2010. Specific increases within global decreases: a functional magnetic resonance imaging investigation of five days of motor sequence learning. *J. Neurosci.* 30:8332–41

Sun R, Merrill E, Peterson T. 2001. From implicit skills to explicit knowledge: a bottom-up model of skill learning. *Cogn. Sci.* 25:203–44

Taatgen NA, Huss D, Dickison D, Anderson JR. 2008. The acquisition of robust and flexible cognitive skills. *J. Exp. Psychol.: Gen.* 137:548–65

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Tappe K, Tarves E, Oltarzewski J, Frum D. 2013. Habit formation among regular exercisers at fitness centers: an exploratory study. *J. Phys. Act. Health* 10:607–13
Thøgersen J. 2012. The importance of timing for breaking commuters' car driving habits. *Collegium* 12:130–40
Thorndike EL. 1898. Animal intelligence: an experimental study of the associative processes in animals. *Psychol. Monogr. Gen. Appl.* 2:1–109
Thrailkill EA, Bouton ME. 2015. Contextual control of instrumental actions and habits. *J. Exp. Psychol.: Anim. Learn. Cogn.* 41:69–80
Tobias R. 2009. Changing behavior by memory aids: a social psychological model of prospective memory and habit development tested with dynamic field data. *Psychol. Rev.* 116:408–38
Tolman EC. 1948. Cognitive maps in rats and men. *Psychol. Rev.* 55:189–208
Triandis HC. 1977. *Interpersonal Behavior*. Monterey, CA: Brooks/Cole Publ.
Tricomi E, Balleine BW, O'Doherty JP. 2009. A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29:2225–32
Valentin VV, Dickinson A, O'Doherty JP. 2007. Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.* 27:4019–26
Verplanken B, Aarts H. 1999. Habit, attitude, and planned behaviour: Is habit an empty construct or an interesting case of goal-directed automaticity? *Eur. Rev. Soc. Psychol.* 10:101–34
Verplanken B, Aarts H, Van Knippenberg A. 1997. Habit, information acquisition, and the process of making travel mode choices. *Eur. J. Soc. Psychol.* 27:539–60
Verplanken B, Orbell S. 2003. Reflections on past behavior: a self-report index of habit strength. *J. Appl. Soc. Psychol.* 33:1313–30
Verplanken B, Walker I, Davis A, Jurasek M. 2008. Context change and travel mode choice: combining the habit discontinuity and self-activation hypotheses. *J. Environ. Psychol.* 28:121–27
Vishwanath A. 2015. Examining the distinct antecedents of e-mail habits and its influence on the outcomes of a phishing attack. *J. Comput. -Mediat. Commun*. doi: 10.1111/jcc4.12126
Vohs KD, Baumeister RF, Ciarocco NJ. 2005. Self-regulation and self-presentation: regulatory resource depletion impairs impression management and effortful self-presentation depletes regulatory resources. *J. Personal. Soc. Psychol.* 88:632–57
Voon V, Derbyshire K, Rück C, Irvine MA, Worbe Y, et al. 2015. Disorders of compulsivity: a common bias towards learning habits. *Mol. Psychiatry* 20:345–52
Walker I, Thomas GO, Verplanken B. 2014. Old habits die hard: travel habit formation and decay during an office relocation. *Environ. Behav.* doi: 10.1177/0013916514549619
Wansink B, Payne CR. 2012. Eating behavior and obesity at Chinese buffets. *Obesity* 16:1957–60
Wason PC, Evans J. 1975. Dual processes in reasoning? *Cognition* 3:141–54
Webb TL, Sheeran P. 2006. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychol. Bull.* 132:249–68
Webb TL, Sheeran P, Luszczynska A. 2009. Planning to break unwanted habits: Habit strength moderates implementation intention effects on behaviour change. *Br. J. Soc. Psychol.* 48:507–23
Wise RA. 2004. Dopamine, learning and motivation. *Nat. Rev. Neurosci.* 5:483–94
Wood W, Labrecque JS, Lin PY, Rünger D. 2014. Habits in dual process models. In *Dual Process Theories of the Social Mind*, ed. JW Sherman, B Gawronski, Y Trope, pp. 371–85. New York: Guilford
Wood W, Neal DT. 2007. A new look at habits and the habit-goal interface. *Psychol. Rev.* 114:843–63
Wood W, Neal DT. 2009. The habitual consumer. *J. Consum. Psychol.* 19:579–92
Wood W, Quinn JM, Kashy DA. 2002. Habits in everyday life: thought, emotion, and action. *J. Personal. Soc. Psychol.* 83:1281–87
Wood W, Tam L, Witt MG. 2005. Changing circumstances, disrupting habits. *J. Personal. Soc. Psychol.* 88:918–33
Wunderlich K, Dayan P, Dolan RJ. 2012. Mapping value based planning and extensively trained choice in the human brain. *Nat. Neurosci.* 15:786–91
Wymbs NF, Grafton ST. 2014. The human motor system supports sequence-specific representations over multiple training-dependent timescales. *Cereb. Cortex*. doi: 10.1093/cercor/bhu144
Yin HH, Knowlton BJ. 2006. The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7:464–76

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Yin HH, Knowlton BJ, Balleine BW. 2004. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19:181–89

Yin HH, Knowlton BJ, Balleine BW. 2006. Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav. Brain Res.* 166:189–96

Yin HH, Ostlund SB, Knowlton BJ, Balleine BW. 2005. The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.* 22:513–23

Zapata A, Minney VL, Shippenberg TS. 2010. Shift from goal-directed to habitual cocaine seeking after prolonged experience in rats. *J. Neurosci.* 30:15457–63

Annu. Rev. Psychol. 2016.67. Downloaded from www.annualreviews.org
Access provided by University of California - Santa Barbara on 09/13/15. For personal use only.

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

The Journal of Neuroscience, March 1993, 13(3): 900–913

Responses of Monkey Dopamine Neurons to Reward and Conditioned Stimuli during Successive Steps of Learning a Delayed Response Task

Wolfram Schultz, Paul Apicella,[a] and Tomas Ljungberg[b]

Institut de Physiologie, Université de Fribourg, CH-1700 Fribourg, Switzerland

The present investigation had two aims: (1) to study responses of dopamine neurons to stimuli with attentional and motivational significance during several steps of learning a behavioral task, and (2) to study the activity of dopamine neurons during the performance of cognitive tasks known to be impaired after lesions of these neurons. Monkeys that had previously learned a simple reaction time task were trained to perform a spatial delayed response task via two intermediate tasks. During the learning of each new task, a total of 25% of 76 dopamine neurons showed phasic responses to the delivery of primary liquid reward, whereas only 9% of 163 neurons responded to this event once task performance was established. This produced an average population response during but not after learning of each task. Reward responses during learning were significantly more numerous and pronounced in area A10, as compared to areas A8 and A9. Dopamine neurons also showed phasic responses to the two conditioned stimuli. These were the instruction cue, which was the first stimulus in each trial and indicated the target of the upcoming arm movement (58% of 76 neurons during and 44% of 163 neurons after learning), and the trigger stimulus, which was a conditioned incentive stimulus predicting reward and eliciting a saccadic eye movement and an arm reaching movement (38% of neurons during and 40% after learning). None of the dopamine neurons showed sustained activity in the delay between the instruction and trigger stimuli that would resemble the activity of neurons in dopamine terminal areas, such as the striatum and frontal cortex. Thus, dopamine neurons respond phasically to alerting external stimuli with behavioral significance whose detection is crucial for learning and performing delayed response tasks. The lack of sustained activity suggests that dopamine neurons do not encode representational processes, such as working memory, expectation of external stimuli or reward, or preparation of movement. Rather, dopamine neurons are involved with transient changes of impulse activity in basic attentional and motivational processes underlying learning and cognitive behavior.

[Key words: behavior, motivation, learning, conditioning, reward, cognition, attention]

In previous neurophysiological studies in behaving monkeys, it was found that dopamine (DA) neurons respond to primary food and fluid rewards and to conditioned incentive stimuli predicting reward (Schultz, 1986; Romo and Schultz, 1990; Ljungberg et al., 1992). Responses to conditioned stimuli have also been observed in awake, haloperidol-treated rats (Miller et al., 1981). DA neurons also respond to novel, unexpected stimuli, both in cats (Steinfels et al., 1983) and in monkeys (Ljungberg et al., 1992). These data suggest that DA neurons respond to salient environmental stimuli that alert the animal and attract its attention. Large, albeit not exclusive, classes of these effective stimuli are primary and learned appetitive motivating stimuli.

Primary rewards, novel stimuli, and conditioned incentive stimuli play particularly prominent roles when an animal learns new or changed behavioral contingencies for obtaining reward. We have investigated the responses of DA neurons to various stimuli while animals learned simple reaction time tasks (Ljungberg et al., 1992). It was found that neurons responded to the sight or reception of primary reward during learning, and that these responses were transferred to the conditioned stimulus predicting reward when task performance was established. These responses to the conditioned stimulus were strongly reduced after overtraining. These data suggest that the responses of DA neurons may be particularly important during learning when the animal adapts its behavior to new situations.

The present study aimed to answer two questions. (1) If the responses of DA neurons are transferred between the most significant stimuli during the course of learning, then how would these neurons respond when the learning of a more elaborate task requires a consecutive series of intermediate subtasks and does not occur in basically a single step, as with a reaction time task? More complicated tasks performed by monkeys in a laboratory setting could comprise the choice of different spatial targets for the behavioral response and involve working memory and the preparation of movement. (2) It would be particularly interesting to see how the activity of DA neurons changes during the performance of such tasks, since Parkinsonian patients and animals with lesions of DA neurons show considerable difficulties in performing tasks requiring the use of declarative cognitive or procedural representations (Brozoski et al., 1979; Simon et al., 1980; Cools et al., 1984; Saint-Cyr et al., 1988; Schneider and Kovelowski, 1990). In combining these two questions, we

Received June 17, 1992; revised Sept. 3, 1992; accepted Sept. 9, 1992.

We thank Dr. J. R. Hollerman for comments on the manuscript and F. Tinguely, A. Schwarz, and J. Corpataux for technical assistance. The study was supported by the Swiss NSF (Grants 3.473-0.86, 31-28591.90) and the United Parkinson Foundation. Postdoctoral fellowships to P.A. were provided by the Fyssen Foundation and the Fondation pour la Recherche Médicale.

Correspondence should be addressed to Dr. W. Schultz at the above address.

[a]Present address: Laboratory of Cellular and Functional Neurobiology, CNRS, F-13274 Marseille, France.

[b]Present address: Department of Pharmacology, Karolinska Institute, S-10401 Stockholm, Sweden.

Copyright © 1993 Society for Neuroscience 0270-6474/93/130900-14$05.00/0

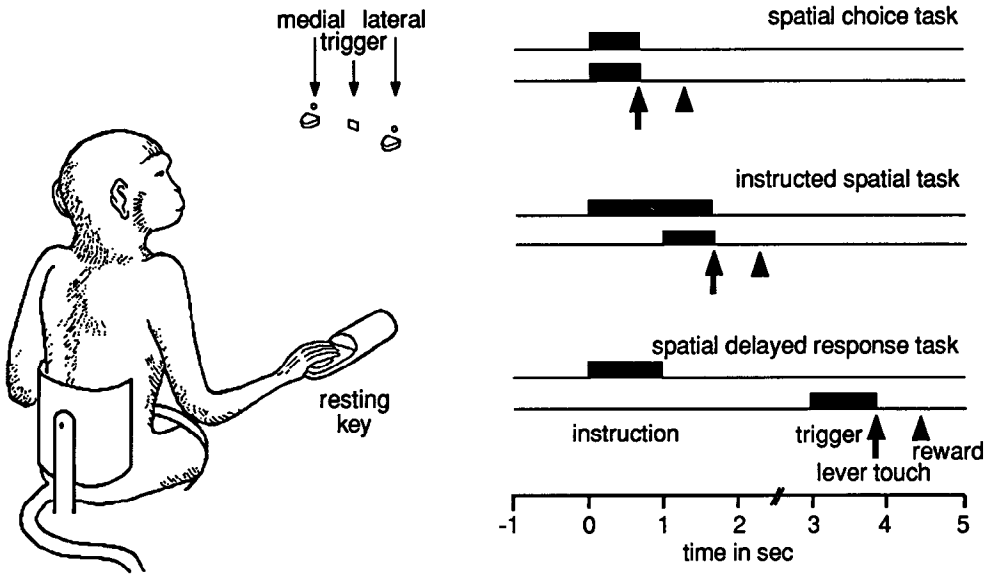


Figure 1. Schematic description of behavioral tasks. A trial is initiated when the animal keeps its hand relaxed on the immovable resting key. Illumination of an instruction light above the medial or lateral lever indicates the future target of reaching. The trigger light determines the time when the resting key should be released and the lever indicated by the instruction touched. Medial and lateral levers are used in random alternation. In the spatial choice task, both instruction and trigger lights come up at the same time and are extinguished upon lever touch. In the instructed spatial task, the trigger light comes up 1 sec after the instruction, and both lights are extinguished upon lever touch. In the spatial delayed response task, the trigger comes up after a randomly varying interval of 2.5–3.5 sec after instruction onset. The instruction is extinguished 1 sec after it came up. Thus, the animal cannot precisely predict when the trigger light comes up and needs to retain the spatial position of the lever until it is allowed to move in response to the trigger. Reward is given in all tasks 0.5 sec after correct lever touch.

studied DA neurons while monkeys learned in several steps a spatial delayed response task in which they needed to remember the spatial position of a cue for several seconds, prepared for an upcoming arm movement, and executed it when a trigger stimulus came up. Because we only wanted the animals to advance in learning the tasks while we recorded from a DA neuron, task acquisition was facilitated by using animals that had already learned a simple reaction time task (Ljungberg et al., 1992), and by employing the most basic version of a delayed response task with only two reaching targets that were constantly visible to the animals.

These data have been presented previously in short form (Schultz et al., 1991).

Materials and Methods

The study was performed on two male *Macaca fascicularis* monkeys of 3.5 kg weight. Activity of single neurons was recorded with moveable microelectrodes during learning and established performance of three consecutive spatial reaching tasks. Animals were killed for histological reconstruction of recording sites. A third monkey was used before the recording experiments for the development of task learning. The behavioral apparatus and recording techniques were similar to those previously reported (Schultz, 1986; Ljungberg et al., 1992).

Behavioral procedures

The behavioral apparatus was positioned at reaching distance (250 mm from the animal's shoulder) in the right half of the frontal wall of a completely enclosed primate chair (Fig. 1). A yellow, rectangular light-emitting diode (11 × 11 mm) was mounted at 27° lateral to the midsagittal plane and at eye level of the animal. Two small levers (7 × 15 mm) were placed 40 mm medial and lateral to the yellow light. Levers protruded by 20 mm from the frontal wall and made electrical contact upon downward movement of 1 mm. One green, round light-emitting diode of 3 mm diameter was located 10 mm above each lever. An immovable, touch-sensitive key was mounted below the lights at waist level such that the elbow joint of the animal was kept at approximately 90° when its hand rested on the key. A drop of apple juice (0.15–0.20 ml) served as reward, which was delivered by an electronically controlled solenoid valve. The solenoid emitted an audible click upon complete opening at 18 msec after the electronic pulse. Liquid arrived at the spout in front of the animal's mouth with an average delay of 55 msec after the electronic pulse. Two closed-circuit video systems served to supervise limb and mouth movements continuously. Monkeys were deprived of fluid during weekdays. They were released into their home cages after each daily experiment of 3–4 hr and received water ad libitum during the subsequent 1 hr.

Both animals had been conditioned to perform in a simple reaction time task (Ljungberg et al., 1992) in which they kept their one hand relaxed on the resting key until the yellow light came on as trigger stimulus. The animal released the key in response to this stimulus, reached out, and depressed the small lever below the light for liquid reward. In the present tasks, a spatial component of reaching was introduced by using two target levers instead of one. The correct lever was indicated by the instruction light shown at different times in the three tasks.

Spatial choice task. The instruction light situated above one of the levers came up at the same time as the central trigger light. The animal released the key in response to the two simultaneous lights, reached out, and depressed the lever indicated by the instruction light in order to receive the reward. Both lights were extinguished upon lever touch or, if no movement occurred, after 1 sec. Liquid reward was delivered 500 msec after correct lever touch. This delay allowed the separation of neuronal responses to lever touch from those to reward. Trials lasted 6–9 sec; intertrial intervals varied between 3 and 5 sec. The use of the lateral and medial levers alternated randomly between trials, but the successive use of the same lever was limited to maximal three trials.

Instructed spatial task. The instruction light situated above one of the levers came up at a fixed interval of 1 sec before the central trigger light. The animal released the resting key and touched the lever in response to the trigger light. Thus, the instruction light indicated the lever to be touched and served as preparatory signal for the upcoming behavioral response, whereas the trigger light determined the time of responding without providing spatial information. Both lights were extinguished upon lever touch or, if no movement occurred, after 1 sec. Reward was delivered 500 msec after correct lever touch.

Spatial delayed response task. The instruction light indicating the lever to be touched came up for a fixed duration of 1 sec. The central light triggering the reaching movement was illuminated at a randomly varied interval of 2.5–3.5 sec after instruction onset and was extinguished upon

lever touch or, if no movement occurred, after 1 sec. Reward was delivered 500 msec after touch of the correct lever. This situation constitutes a spatial delayed response task in which the animal has to remember the lever to be touched.

Learning of tasks. The animals, previously trained in the simple reaction time task, were first conditioned to perform the spatial choice task. After reaching stable performance for about 1 week, the task was abandoned and the instructed spatial task was introduced. After stable performance for about 1 week, this task was in turn abandoned for the spatial delayed response task. During conditioning in the spatial choice and the instructed spatial tasks, the new contingencies were introduced in one step and animals had to find the correct solution to the new problem by their own trial and error attempts. Learning of the spatial delayed response task involved a stepwise approach to the longer interval and delay. For monkey A, intervals between onsets of instruction and trigger stimuli were gradually prolonged to 3 sec, and the duration of the instruction cue was subsequently shortened to 1 sec, thus introducing the delay. For monkey B, the duration of the instruction cue was kept at 1 sec, and intervals between instruction and trigger stimuli were gradually prolonged. The acceptable interval for lever pressing after the trigger stimulus was initially set to 3 sec and gradually shortened to the final 1 sec.

Learning of each task was best described by the frequency of correct responses. Systematic changes in arm reaction time and movement time were not observed during learning, possibly because of extensive training in the previous reaction time task involving a comparable arm movement. Thus, each task was divided into the learning and the established performance parts according to the time when a stable frequency of correct responses was reached.

Data acquisition

All behavioral performance was controlled by a suitably interfaced laboratory computer, the selection of appropriate task parameters being done by the experimenters individually for each block of trials. The different lights and the solenoid for reward delivery were driven by output pulses from the computer. Key release was detected by a frequency-sensing circuit that reacted to a change in electrical capacity induced by the touch of the animal's hand. Errors in behavioral performance led either to cancellation of all further signals in a given trial, including reward, or to immediate trial restart.

Animals had been implanted for the preceding study on reaction time tasks. Under deep sodium pentobarbital anesthesia and aseptic conditions, cylinders for head fixation and a stereotaxically positioned, stainless steel chamber were fixed to the skull to permit vertical access with microelectrodes to the left substantia nigra. The dura was left intact. Teflon-coated, multistranded, stainless steel wires were implanted into the extensor digitorum communis and biceps muscles of the arm and led subcutaneously to the head. Ag-AgCl electrodes were implanted into the outer, upper, and lower canthi of the orbits. All metal components, including connector plugs for the muscle and periorbital electrodes, were embedded in several layers of dental cement and fixed to the skull with surgical grade stainless steel screws. One week after implantation, animals were anesthetized with pentobarbital and the area of substantia nigra was localized by taking lateral and coronal radiographs with a guide cannula installed at a known coordinate in reference to the implanted steel chamber. The ventroposteromedial thalamus overlying the lateral substantia nigra was electrophysiologically explored for trigeminal input on the same occasion, and later occasionally in the waking animal.

The activity of single neurons was recorded extracellularly with glass-insulated, platinum-plated tungsten microelectrodes (exposed tips of 9–16 μm length and 2.5–3.5 μm diameter), which were passed each day into the brain, together with and inside of a rigid guide cannula of 0.6 mm outer diameter. Microelectrodes were moved vertically in the stereotaxic plane in parallel tracks that conformed to a 1 mm grid. Signals from the microelectrode were conventionally amplified, filtered (100 Hz lower cutoff), and monitored with oscilloscopes and audiomonitors using earphones to prevent the monkeys from hearing the neuronal activity. Full waveforms of impulses from each neuron were displayed on a digital oscilloscope using the pretrigger viewing facility and subsequently stored on computer disks. Somatodendritic discharges were discriminated against those originating from fibers using earlier established criteria, in particular the very short durations of fiber impulses (0.1–0.3 msec) (Hellweg et al., 1977; Schultz and Romo, 1987). Neuronal discharges were also converted into standard digital pulses by means of an adjustable Schmitt trigger, the output of which was continuously monitored on the digital oscilloscope together with the original waveform. DA neurons were searched for in a systematic fashion. Every neuron fulfilling the criteria for being dopaminergic was tested with at least 20 trials of a given task situation and included in the study.

Electromyograms (EMGs) were collected during all neuronal recordings through the chronically implanted electrodes. EMG activity was filtered (10–250 Hz bandpass; −12 dB at 1 kHz), rectified, and displayed on conventional oscilloscopes. Rectified EMG activity was also passed through an adjustable Schmitt trigger and fed to the computer. Horizontal and vertical electrooculograms (EOGs) were collected during all neuronal recordings from the implanted periorbital electrodes. The gain of ocular electrodes and positions of the eyes were calibrated by having the animal fixate small morsels of preferred food presented at several known horizontal and vertical eccentricities while the frontal enclosure of the primate chair was kept open. Direct current offset had to be adjusted once every 3–4 weeks.

All behavior-related digital signals and pulses from neuronal discharges and EMG activity were sampled on line as bits in parallel at a rate of 2 kHz by a laboratory computer. Analog signals from EOGs and rectified EMGs were sampled after analog-to-digital conversion at a rate of 2 kHz by the computer. Eight consecutive analog values were averaged to obtain a final temporal resolution of 4 msec (0.25 kHz) for data storage. The behavioral relationships of neuronal discharges, digital and analog EMG activity, and EOGs were displayed in each trial on line on the computer video screen in the form of dot displays and analog curves. All data were stored uncondensed on computer disks.

Data analysis

Because of the stereotyped character of phasic responses of DA neurons to external stimuli seen before (Schultz, 1986; Romo and Schultz, 1990; Schultz and Romo, 1990), a standard time window procedure was developed in order to quantify and compare objectively the neuronal responses between different parts of experimentation (Ljungberg et al., 1992). First, onset and offset times of all phasic activations in response to the conditioned stimuli and delivery of liquid reward were determined with a resolution of 4 msec from the first or last of at least three consecutive bins in which activity was above or back to baseline, respectively, as indicated by dot displays, perievent time histograms, and the slopes of cumulative frequency distributions of neuronal impulses. The significance of an increase in activity over baseline was assessed with a specially implemented procedure using the two-tailed Wilcoxon matched-pairs signed-rank test ($p < 0.01$; Schultz, 1986). Depressant responses to stimuli were tested for statistical significance of change in the standard time window, whereas onset and offset times were not measured in view of the low baseline activity of DA neurons.

For subsequent analysis, the standard time window was set to include 80% of onset and offset times of statistically significant increases in activity. Obtained time windows were 88–200 msec after onsets of instruction and trigger lights and 172–256 msec after delivery of liquid reward. Two measures for the responsiveness of neurons were sought. (1) The number of individual neurons showing a significant increase in activity in each time window was obtained with the Wilcoxon procedure. (2) The responsiveness of the whole population of neurons was assessed by measuring the magnitude of change of activity in the time windows against baseline activity for *every* neuron recorded in a particular task situation, independent of a significant response. For both measures, baseline activity was determined during the 500 msec interval preceding presentation of the first stimulus of each trial. All neurons used for determining onset and offset times of significant increases in activity also showed significant activations in the standard time window. Magnitudes determined during the individual durations of response of each *responding* neuron were significantly higher by 20–30% than in the standard time windows for instruction, trigger, and reward used with the same neurons during both learning and established task performance in all three tasks ($p < 0.001$, paired t test).

In addition to the quantitative assessment of responses of individual neurons, the overall responsiveness of the entire population of neurons during particular tasks and learning phases was visually assessed from averages of neuronal activity. A peristimulus time histogram was constructed for every neuron recorded independent of a significant response, and the content of each of its bins was divided by the number of trials to obtain a normalized count. All normalized histograms referenced to the same behavioral event were averaged to obtain the population histogram.

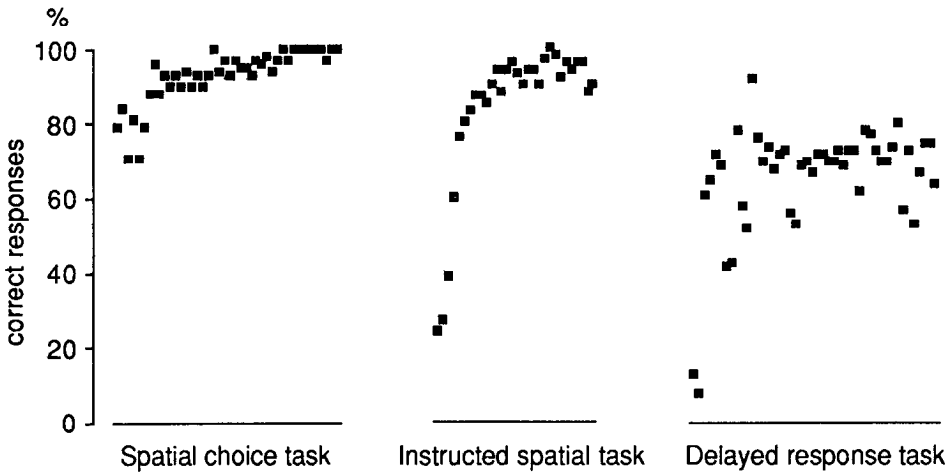


Figure 2. Learning curves for the three tasks in one monkey. Each data point was obtained during the recording of one DA neuron. Tasks were successively employed in the order shown from left to right.

Behavioral performance was assessed during the recording period of every neuron by calculating the median (50th percentile) of reaction time (from trigger light to key release) and movement time (from key release to lever touch) of correct behavioral responses.

Response magnitudes and medians of behavioral parameters were normally distributed (Kolmogorov–Smirnov one-sample test, $p > 0.05$). Parametric statistics were therefore used, which included the means, standard errors of the mean (SEMs), and two-tailed Student's t test for one sample and for paired and unpaired two samples. All timing and magnitude values are given as means ± SEM. Correlations between magnitudes of changes after different stimuli were determined with Spearman's rank correlation coefficient. The χ^2 test was employed for comparing the distributions of fractions of responding neurons among groups A8, A9, and A10 against a hypothetically homogeneous distribution. Neuronal changes obtained from the standard time window method were compared with one-way and two-way ANOVAs using monkeys, phases of experimentation, or midbrain areas as factors. Data from neuronal activity were pooled over the two monkeys when systematic differences were absent.

Histological reconstruction

During the last recording sessions with each animal, small marking lesions were placed by passing negative currents (10–20 μA for 10–20 sec) through the microelectrode immediately after recording from a neuron in the substantia nigra, while larger lesions (20 μA for 20 or 60 sec) were positioned at a few locations above in the same track. This produced distinct patterns of vertically oriented histological marks. Animals were deeply anesthetized with sodium pentobarbital and conventionally perfused with formaldehyde through the heart. Guide cannulas were inserted into the brain at known coordinates of the implant system in order to delineate the general area of recording. The tissue was cut in 50-μm-thick serial coronal sections on a cryotome and stained with cresyl violet. All histological sections were projected on paper, and the outlines of brain structures and the marks from lesions and recent electrode tracks were drawn. Recording positions in tracks marked by electrolytic lesions were reconstructed by using the distances to the lesions according to protocolled micrometer readings. Positions in parallel neighboring tracks were reconstructed at comparable vertical levels. The recorded DA neurons were attributed to groups A8, A9, and A10 according to previous histological descriptions of catecholamine cell groups in the monkey midbrain (Felten and Sladeck, 1983). Thus, group A9 comprised the pars compacta of substantia nigra, whereas group A8 was located dorsolaterally to the substantia nigra in the lateral reticular formation, and group A10 was found dorsomedially adjoining to and incompletely separated from the medial substantia nigra (lateral A10 region of Felten and Sladeck, 1983).

Results

General

Behavior. The gradual learning of each task is illustrated by the learning curves of Figure 2. Errors in the instructed spatial task consisted in precocious release of the resting key before the trigger stimulus, suggesting that animals used the instruction as stimulus for an appropriately delayed movement response. This was probably due to previous overtraining in the reaction time task in which animals moved in response to the first and only stimulus of each trial. In the spatial delayed response task, the interval between instruction and trigger stimuli was prolonged and varied to such an extent that animals only moved in response to the trigger light. Errors in this task mostly consisted in touching the wrong lever.

Performance in the spatial choice task levelled off at almost 100% correct responses, whereas it was successively lower in the two consecutive tasks of higher complexity. Learning of each task was considered to be terminated with a level of performance of 95%, 90%, and 75% correct responses reached in the majority of blocks of trials in spatial choice, instructed spatial, and spatial delayed response tasks, respectively. During established task performance, reaction and movement times in the spatial choice and the instructed spatial tasks were in the same range as in the simple reaction time task employing the same movement (Ljungberg et al., 1992), whereas reaction times were considerably longer in the spatial delayed response task ($p < 0.01$, Fisher test following one-way ANOVA) (Table 1).

Neurons. A total of 239 DA neurons were recorded in group A9 (pars compacta of substantia nigra, $n = 162$), A8 ($n = 38$), and A10 ($n = 39$). Neurons were recorded during learning and established performance of the spatial choice task, followed by recordings during learning and established performance of the instructed spatial task, and subsequently the spatial delayed response task. Histological reconstructions of recording sites revealed unsystematic, statistically insignificant differences in recording areas between tasks, and between learning and established performance of each task (Table 2).

As described before, midbrain DA neurons displayed specific electrophysiological characteristics (Schultz, 1986; Schultz and Romo, 1987). They discharged initially negative or positive impulses with low frequencies (0.5–9 impulses/sec), polyphasic waveforms, and relatively long durations (1.8–5.0 msec). They contrasted with pars reticulata neurons discharging impulses of <1.1 msec duration at rates of 70–90 impulses/sec, a few unknown neurons discharging impulses of <1.0 msec at low rates, and presumptive fibers discharging very short impulses (0.1–0.3 msec).

Table 1. Behavioral performance after established task performance

	Reaction time (msec)	Movement time (msec)	Blocks of trials (*n*)
Spatial choice task			
Monkey A	250 ± 3[1,2]	349 ± 4[1,2]	23
Monkey B	271 ± 3[2]	303 ± 2[2]	32
Instructed spatial task			
Monkey A	224 ± 7[3]	302 ± 5	19
Monkey B	272 ± 7[3]	312 ± 3[3]	16
Spatial delayed response task			
Monkey A	335 ± 3	310 ± 3	31
Monkey B	346 ± 7	277 ± 2	42

Values are given as means ± SEM of performance in individual blocks of trials, each of which contained 40–64 trials. The differences between both monkeys and the three tasks were statistically significant ($p < 0.001$, two-way ANOVA with monkeys and tasks as factors). Separately for each monkey, the following reaction and movement times varied significantly between the three tasks ($p < 0.01$ in Fisher test after one-way ANOVA):

[1], spatial choice versus instructed spatial task;

[2], spatial choice versus spatial delayed response task;

[3], instructed spatial versus spatial delayed response task.

Data from movements to medial and lateral lever were pooled. All measures were obtained in trials in which DA neurons were recorded.

Responses to reward

Of 76 DA neurons recorded during the learning phases of the three tasks, 19 neurons (25%) responded to the delivery of liquid reward. By contrast, after task performance was established only 14 of 163 neurons (9%) were activated by reward (Fig. 3, Table 3). The activity of one neuron was depressed by reward. There were only insignificant differences in reward responses after reaching to the medial versus lateral levers. Of the total of 33 neurons responding to reward, 24 neurons (73%) were also activated by the instruction cue.

The delayed delivery of reward after correct performance showed that reward responses occurred to the delivery of liquid and not to lever touch (Fig. 4, left). The reward response was equally present when reward was delivered simultaneously with correct lever touch. Nine neurons activated by reward in correct trials (235 ± 34%) were also recorded in >10 trials that were

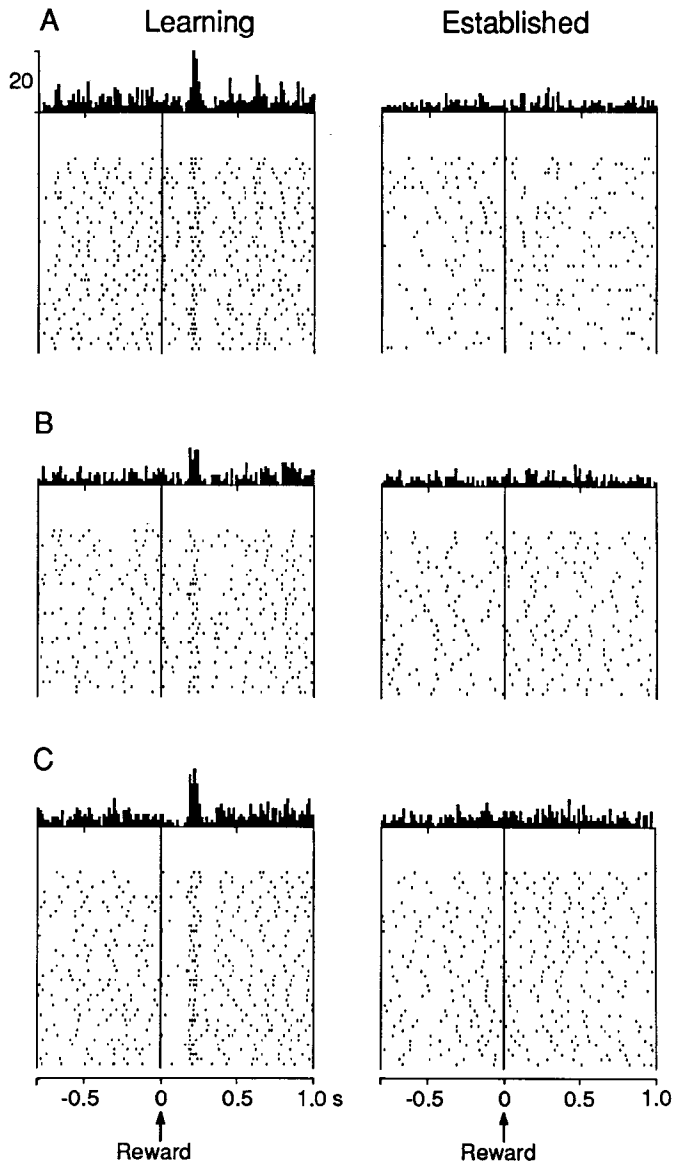


Figure 3. DA neurons respond to reward during learning but not during established task performance. *A*, Spatial choice task. *B*, Instructed spatial task. *C*, Spatial delayed response task. Neuronal activity is aligned on the electronic pulse that drives the solenoid valve delivering the reward liquid. Each panel shows the perievent time histogram and rasters of impulses from one neuron. Distances of *dots* to reward correspond to real-time intervals. Each *line of dots* shows one trial. Trials are pooled over medial and lateral levers that alternated randomly as targets for reaching during the experiment. Neurons shown on the left and right, respectively, belonged to the following groups: *A*, both A9; *B*, A10 and A9; *C*, both A10. Vertical scale is 20 impulses/bin.

Table 2. Regional distribution of neurons recorded in the two monkeys

	Task learning	Established performance
Group		
A8	15 (20%)	23 (14%)
A9	42 (55%)	120 (74%)
A10	19 (25%)	20 (12%)
Midbrain DA zone[a]		
Lateral	6 (8%)	9 (6%)
Intermediate	50 (66%)	125 (76%)
Medial	20 (26%)	29 (18%)
Sum	76 (100%)	163 (100%)

[a] The area of midbrain DA neurons was divided into three equal mediolateral zones.

Table 3. Numbers of neurons responding to reward

	Learning	Established
Spatial choice task	8 of 26 (31%)	2 of 55 (4%)
Instructed spatial task	5 of 22 (23%)	4 of 35 (11%)
Spatial delayed response task	6 of 28 (21%)	8 of 73 (11%)
Sum	19 of 76 (25%)	14 of 163 (9%)

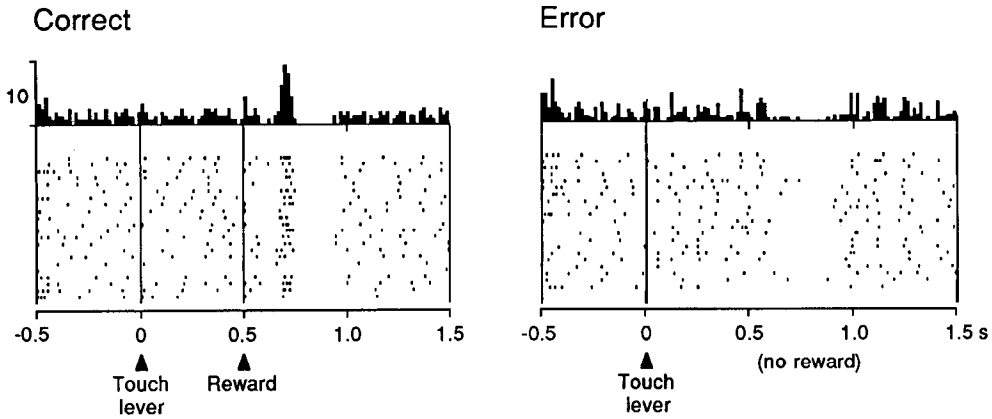


Figure 4. Activating response to reward is replaced by pure depression in erroneous trials at the time of expected reward. *Left*, Reward is delivered 500 msec after correct lever touch. Neuronal activity is aligned on lever touch. *Right*, Both reward delivery and solenoid noise are absent when the wrong lever is touched. Thus, the last stimulus precedes the depression by >0.5 sec. Left and right panels show activity from the same A10 DA neuron during learning of the spatial delayed response task. Vertical scale is 10 impulses/bin.

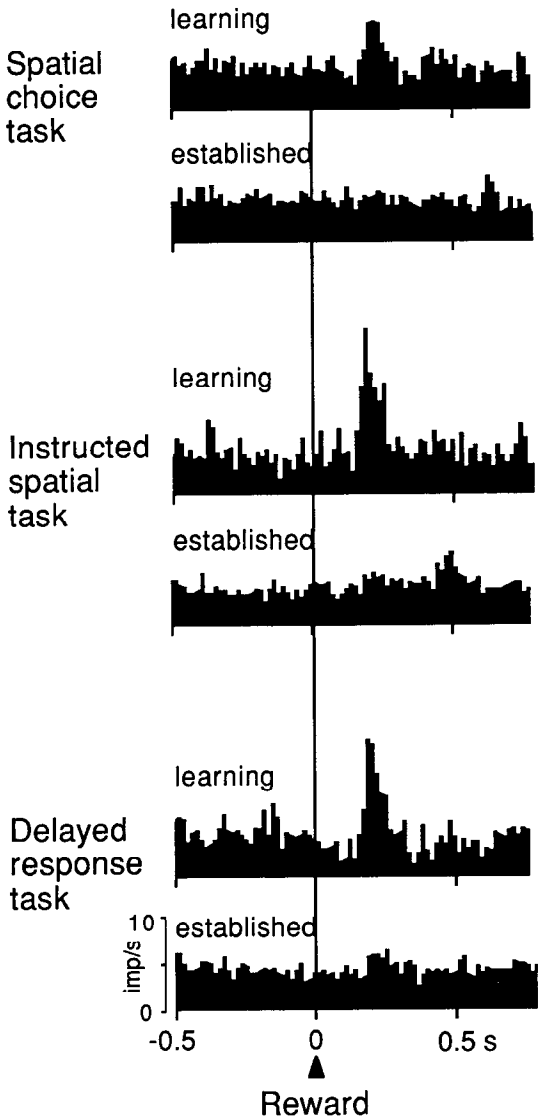


Figure 5. Population responses of DA neurons to reward during learning and established performance of the three tasks. Histograms from each neuron normalized for trial number were added and the resulting sum was divided by the number of neurons. The following data from one monkey were included for spatial choice, instructed spatial, and spatial delayed response tasks, respectively: learning, 853 trials (n = 18 neurons), 264 trials (n = 11), 228 trials (n = 10); established, 1189 trials (n = 23), 803 trials (n = 19), 1198 trials (n = 35). Data were taken from all neurons recorded during each task, were only obtained from correct trials, and were pooled over trials using the medial and lateral levers. Vertical scale denotes impulses/sec.

unrewarded because the incorrect lever was touched. All of them showed depressions of activity at the time at which reward would have been delivered after correct lever touch (mean ± SEM, −59 ± 9%; $p < 0.001$ against 0%, one-sample t test) (Fig. 4, right). Differences between activations in correct and depressions in incorrect trials were significant ($p < 0.001$, paired t test). Thus, the activating reward responses were replaced in error trials by pure depressions at the time of expected reward.

Reward responses in individual neurons showed mean onset latencies of 183 ± 3 msec and lasted for 71 ± 3 msec. There were no significant differences in latency and duration between tasks ($p > 0.01$, one-way ANOVAs). The narrow range and variation of onset latencies and durations suggest a homogeneous temporal response character of DA neurons and allowed us to calculate the average response of all DA neurons studied in each task. This shows that reward responses during but not after learning were sufficiently strong to result in a net response of the whole population (Fig. 5). A similar difference in responsiveness to reward between learning and established performance was found when neuronal activity in the standard time window after reward was compared. Magnitudes of changes in the total neuronal population were significantly higher during learning of all three tasks, as compared to established performance (spatial choice task: 63 ± 14%, n = 26, vs 21 ± 8%, n = 55; instructed spatial task: 106 ± 28%, n = 22, vs 20 ± 7%, n = 35; spatial delayed response task: 60 ± 19%, n = 28, vs 26 ± 9%, n = 73; all $p < 0.001$, t test). Response magnitudes in the time window determined individually for each responding neuron were 242 ± 21% (n = 33). With this 3.5-fold increase, the activity of responding DA neurons increased from about 4.5 impulses/sec baseline to an average of 15 impulses/sec during the response.

Positions of neurons responding to reward in the three tasks are shown in Figure 6. The responsiveness of DA neurons to reward during learning was significantly higher in group A10 than in groups A8 and A9, both in terms of fractions of neurons responding ($p < 0.0001$ in χ^2 test) (Fig. 7) and in magnitudes of changes in the standard time window after reward ($p < 0.01$ in one-way ANOVA). However, the responsiveness during es-

←

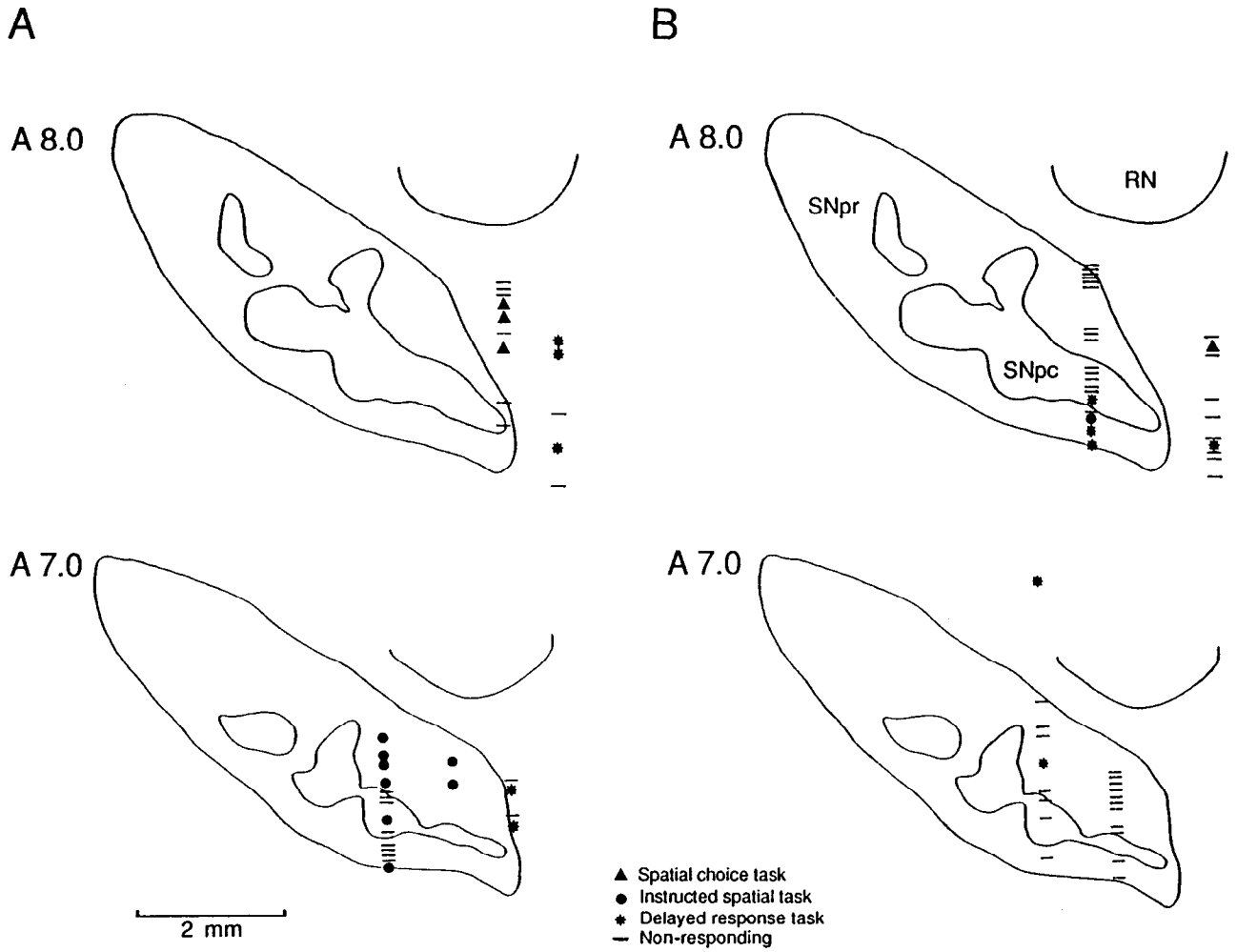


Figure 6. Histological reconstruction of DA neurons responding to reward in one monkey during learning (*A*) and performance (*B*) of the three tasks. Different symbols show positions of neurons responding in the spatial choice task (*triangles*), instructed spatial task (*circles*), and the delayed response task (*asterisks*), whereas *dashes* indicate nonresponding neurons commonly for all tasks. Coronal sections from the left hemisphere are labeled in two coronal stereotaxic planes according to the distances from the interaural line (A7.0 and A8.0). *SNpr* and *SNpc*, pars reticulata and pars compacta of substantia nigra; *RN*, red nucleus.

tablished task performance differed insignificantly in both of these parameters between the three groups ($p > 0.1$).

Responses to conditioned stimuli

Established task performance. The responses of DA neurons to instruction and trigger stimuli differed systematically between the three tasks (Fig. 8, Table 4). In the spatial choice task, in which the instruction and the trigger came up simultaneously, activating responses to the combined instruction–trigger stimulus occurred in 26 of 55 DA neurons (47%) (Fig. 8, top). Only two neurons responded with depression of activity.

In the instructed spatial task, in which the trigger stimulus followed the instruction cue with a fixed interval of 1.0 sec, 17 of 35 DA neurons (49%) responded to the instruction cue, whereas only 3 of the 35 neurons (9%) were activated by the trigger stimulus (Fig. 8, middle). Two neurons were depressed by each stimulus. The temporal aspect of the animals' behavior in this task was mainly determined by the instruction cue. This is suggested by the saccadic reaction to the instruction and its absence to the trigger, and by some of the arm movement reaction times that were too short to be related to the trigger stimulus (Fig. 8, middle).

In the spatial delayed response task, the trigger stimulus followed instruction onset with a randomly varying interval of 2.5–3.5 sec, and the instruction cue was only shown for 1 sec. Neuronal responses occurred to both instruction and trigger stimuli and were accompanied by separate ocular reactions to each stimulus (Fig. 8, bottom). Of 73 neurons tested, 28 (38%) showed activating responses to the instruction and 36 (49%) to the trigger, 25 of them being activated by both stimuli. One neuron was depressed by the trigger. The relatedness between instruction and trigger responses is further suggested by the good correlation of magnitudes of changes in the standard time window between instruction and trigger stimuli ($r = 0.8$; $p < 0.01$; Spearman rank correlation test), which contrasts with the weak correlation between the instruction and reward ($r = 0.2$; $p > 0.01$) (Fig. 9).

None of the neurons showed responses that were specifically related to the medial or lateral lever being used (Fig. 10). Magnitudes of responses to the instruction or trigger stimuli in all

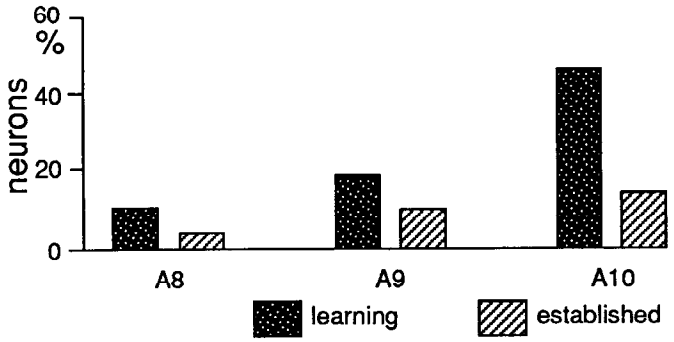
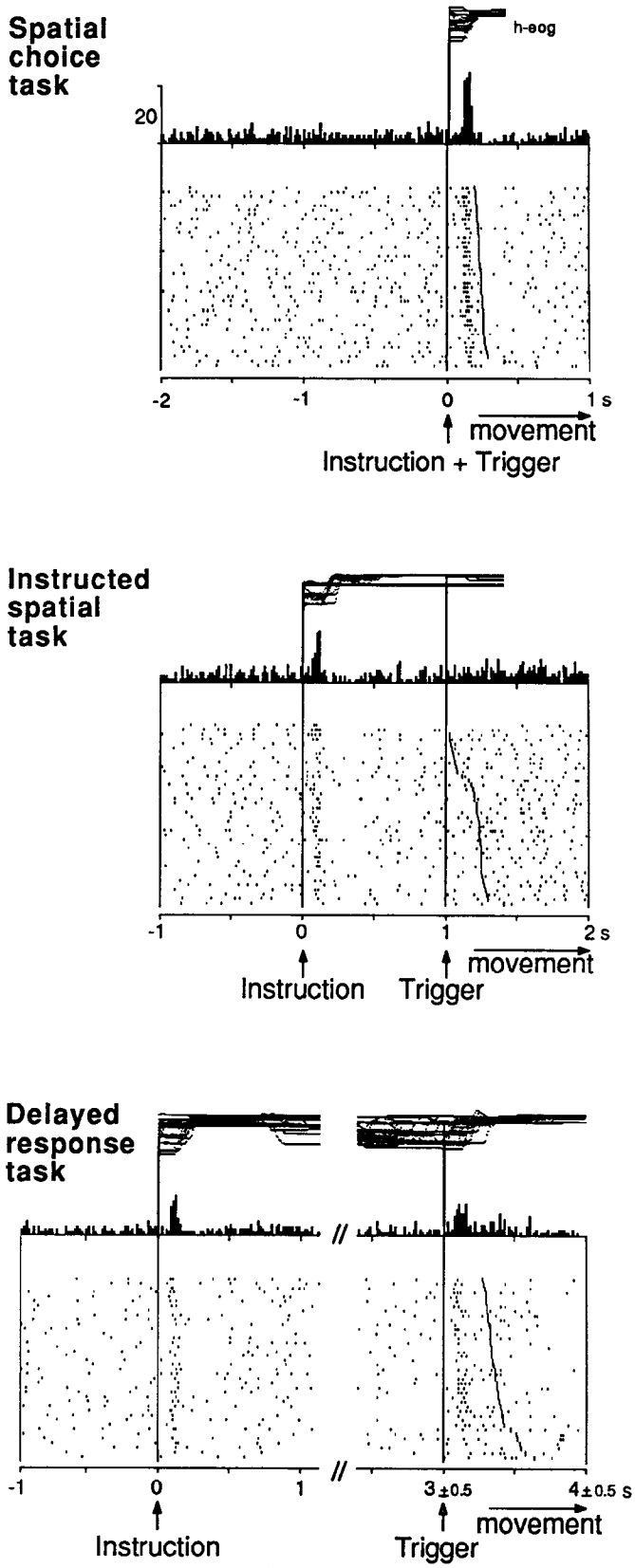


Figure 7. Distribution of DA neurons responding to reward over midbrain cell groups A8–A10 in both monkeys. Neurons from all three tasks are included. *Ordinate* indicates percentage of responding neurons in each cell group. The fraction of neurons responding during learning is significantly higher in group A10, as compared to groups A8 and A9 ($p < 0.0001$, χ^2 test). See Table 2 for numbers of neurons tested.

neurons varied insignificantly in all three tasks between trials using the medial versus the lateral lever.

Responses to instruction and trigger stimuli in the three tasks showed mean onset latencies of 103–112 msec and 101–120 msec, respectively. There were no significant differences between tasks and stimuli ($p > 0.01$, two-way ANOVA), their common mean being 105 ± 2 msec. Durations of instruction responses differed insignificantly between tasks and showed a common mean of 91 ± 3 msec ($p > 0.01$, one-way ANOVA). Durations of trigger responses differed between tasks, being shortest in the instructed spatial task (70 ± 11 msec) and longest in the spatial delayed response task (113 ± 5 msec) ($p < 0.001$, Fisher test after one-way ANOVA). The average response of all neurons recorded in each task illustrates the strong population response to the combined instruction–trigger stimulus during established performance of the spatial choice task, and the slightly weaker responses to individual instruction and trigger stimuli in the two subsequent tasks (Fig. 11). This is also shown by the magnitudes calculated in the standard time window on all neurons recorded. Average magnitudes of changes after the instruction were 126 ± 25%, 75 ± 17%, and 77 ± 12% in the spatial choice, instructed spatial, and spatial delayed response tasks, respectively. Comparable changes after the trigger stimulus were 4 ± 8% and 106 ± 15% in the instructed spatial and spatial delayed response tasks, respectively. Response magnitudes in the time window determined individually for each responding neuron in the three tasks were 277–410% ($n = 71$) and 268–410% ($n = 65$) for instruction and trigger responses, respectively. With this four- to fivefold increase, the activity of responding DA neurons increased from about 4.5 impulses/sec baseline to an average of 17–23 impulses/sec during the response.

Learning. In the spatial choice and instructed spatial tasks, the responsiveness of DA neurons to instruction and trigger stimuli during learning was higher compared to established task



→

Figure 8. Responses of three DA neurons of group A9 to conditioned stimuli during task performance. Superimposed horizontal components of eye movements (*h-eog*) recorded simultaneously with neurons are shown above histograms and rasters of neuronal impulses and illustrate the animal's ocular reaction to stimulus presentation (*up* = eyes to the right). *Small vertical bars* in rasters indicate movement onset (release of resting key). Trials are pooled over medial and lateral levers that alternated randomly as targets for reaching in each block of trials. Trials were ordered off line according to reaction time (from trigger to movement onset). Because of randomly varying intervals between instruction and trigger stimuli (3 ± 0.5 sec), the time base in *C* is split without maintaining the original instruction–trigger interval (rasters separately aligned to instruction and trigger stimuli). Vertical scale is 20 impulses/bin.

Table 4. Numbers of neurons responding to instruction and trigger stimuli

	Learning	Established
Instruction cue		
Spatial choice task	20 of 26 (77%)	26 of 55 (47%)
Instructed spatial task	16 of 22 (73%)	17 of 35 (49%)
Spatial delayed response task	8 of 28 (29%)	28 of 73 (38%)
Sum	44 of 76 (58%)	71 of 163 (44%)
Trigger stimulus		
Spatial choice task	20 of 26 (77%)	26 of 55 (47%)
Instructed spatial task	3 of 22 (14%)	3 of 35 (9%)
Spatial delayed response task	6 of 28 (21%)	36 of 73 (49%)
Sum	29 of 76 (38%)	65 of 163 (40%)

performance, both in terms of fractions of neurons responding and in magnitudes of changes in the standard time windows (Table 4, Fig. 12). In the spatial delayed response task, differences between learning and established performance varied between monkeys, resulting in an overall lower responsiveness during learning.

Recording positions. Positions of neurons recorded from one monkey during the three tasks are shown in Figure 13. During learning of the three tasks, a significantly higher fraction of neurons in group A10 responded to instruction or trigger stimuli, as compared to groups A8 and A9 (both $p < 0.0003$, χ^2 test), although the magnitudes of changes in all neurons tested differed only insignificantly between these three areas ($p > 0.01$, one-way ANOVA) (Fig. 14). Only insignificant regional variations in both the fraction of responding neurons and the magnitudes of changes in all neurons tested were seen during established task performance, for both instruction and trigger stimuli ($p > 0.01$ in both χ^2 and one-way ANOVA tests).

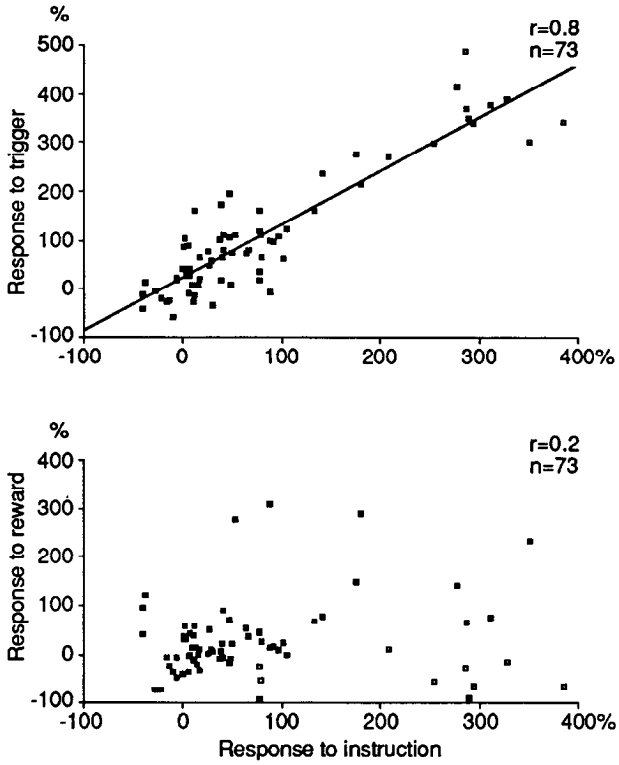


Figure 9. Correlations of magnitudes of neuronal changes between different stimuli. Spearman's rank correlation shows a high coefficient between instruction and trigger stimuli ($r = 0.8$; *top*), but a low coefficient between instruction and reward ($r = 0.2$; *bottom*). Magnitudes of changes in respective standard time windows are shown from all neurons recorded during established task performance in the spatial delayed response task in both monkeys. Mean changes were 77 ± 12%, 106 ± 15%, and 26 ± 9% after instruction, trigger, and reward, respectively.

Discussion

These data show that DA neurons respond to those stimuli that are crucial for learning and performing a behavioral task. In the spatial tasks studied presently, the stimuli effective for activating DA neurons were (1) the delivery of primary reward during learning, which alerts the animal, serves as reinforcer, and indicates correct performance in each task; (2) the instruction cue as first stimulus in each trial, which also provides the necessary spatial information for task performance; (3) the trigger stimulus constituting a conditioned incentive stimulus that predicts reward and elicits an ocular reaction and an arm reaching movement. Most responding neurons were activated by more than one stimulus, indicating that the responses were not selective for the particular behavioral processes associated with the stimuli, such as reinforcement, spatial information, or initiation of movement. Rather, the responses appear to be related to the significance of the stimuli in the particular behavioral situation.

Reward responses during learning

Primary rewards are food objects or fluids that are approached by subjects through innate or instinctive behavior or are learned very early during ontogenesis. Primary rewards may serve to establish and sustain learned behavior, in which case they are called primary reinforcers. A fraction of DA neurons responded to the delivery of primary liquid reward during the learning phase of each of the three successive tasks. The responses were sufficiently frequent and strong to result in a net increase of activity in the whole population of neurons recorded during learning. By contrast, the small fraction of neurons responding to reward during established task performance and their low response magnitude failed to result in such a net population response, suggesting that these responses were little more than random fluctuations in responsiveness among individual neurons. Reward responses during learning occurred more frequently but not exclusively in group A10, suggesting a moderate gradient of responsiveness over groups A8–A10. This might induce a predominant reward-related dopaminergic influence on the postsynaptic structures innervated by A10 neurons, such as the ventral striatum, amygdala, and certain areas of frontal cortex (Szabo, 1980; Porrino and Goldman-Rakic, 1982).

The predominant occurrence of reward responses during the learning phase of each task corresponds to the restriction of similar reward responses to the learning phase of a simple reaction time task (Ljungberg et al., 1992). However, the fraction of responding neurons was only about one-half of that seen during learning of the reaction time task. The present results were obtained in animals trained before in the reaction time task using the same behavioral apparatus (resting key, target lever, mode of reward delivery). It is conceivable that the delivery of reward constitutes a more alerting stimulus for an animal during learning the first task, as compared to learning

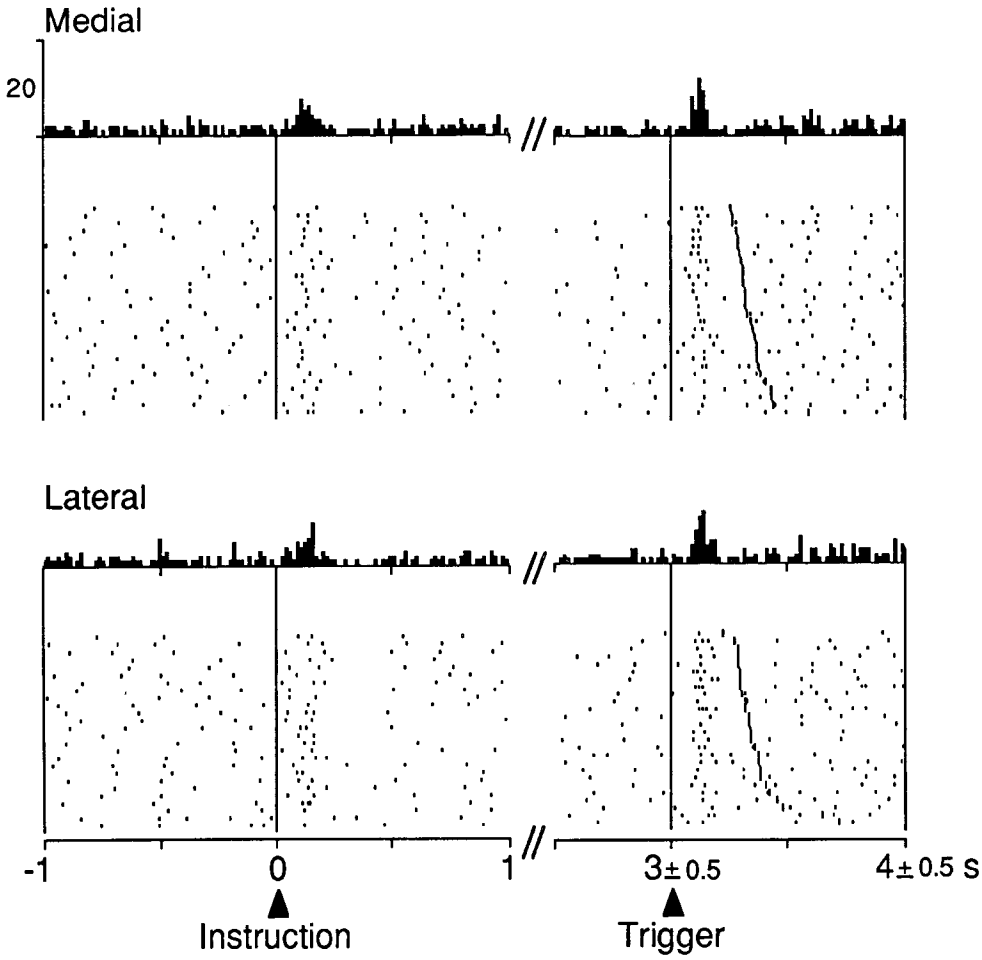


Figure 10. Responses to instruction and trigger stimuli are independent of movement target. Data are from one A9 DA neuron during established performance in the delayed response task. Trials using medial and lateral levers alternated randomly during each block and were separated for analysis and ordered according to reaction time (from trigger to movement onset). *Small vertical lines* in rasters indicate movement onset (release of resting key). Because of randomly varying intervals between instruction and trigger stimuli (3 ± 0.5 sec), the time base is split without maintaining the original instruction–trigger interval. Vertical scale is impulses/bin.

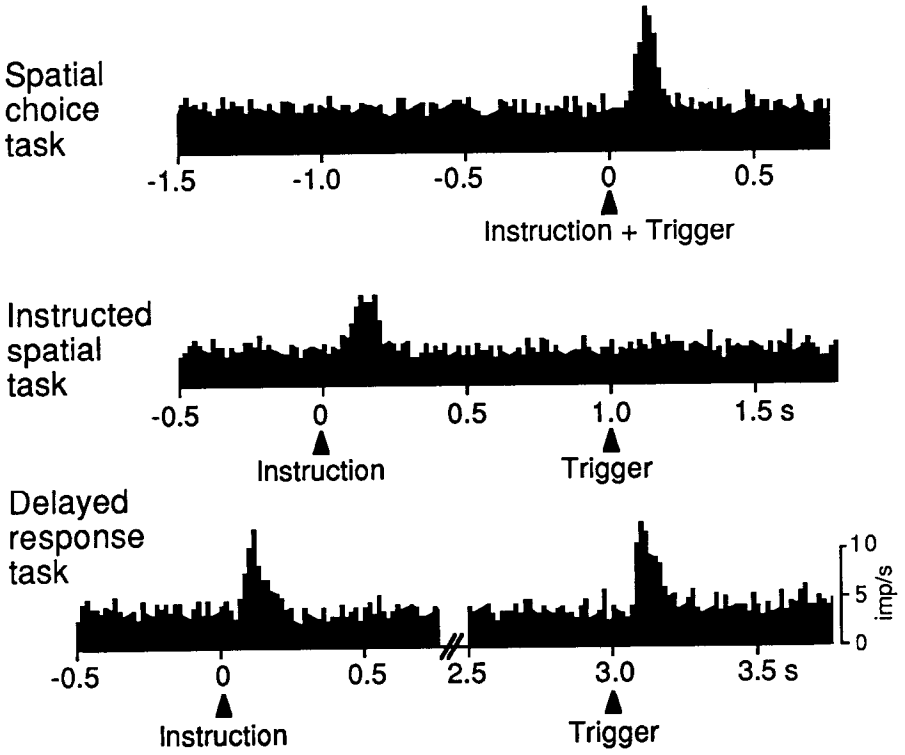


Figure 11. Population responses of DA neurons to instruction and trigger stimuli during established performance of the three tasks. Because of varying intervals, histograms were separately referenced to instruction onset and trigger stimulus for the spatial delayed response task. The following data were included from one monkey for spatial choice, instructed spatial, and spatial delayed response tasks, respectively: 1189 trials ($n = 23$ neurons), 803 trials ($n = 19$), 1198 trials ($n = 35$). Data were taken from all neurons recorded during each task and pooled over correctly performed trials using the medial and lateral levers. Vertical scale is impulses/bin.

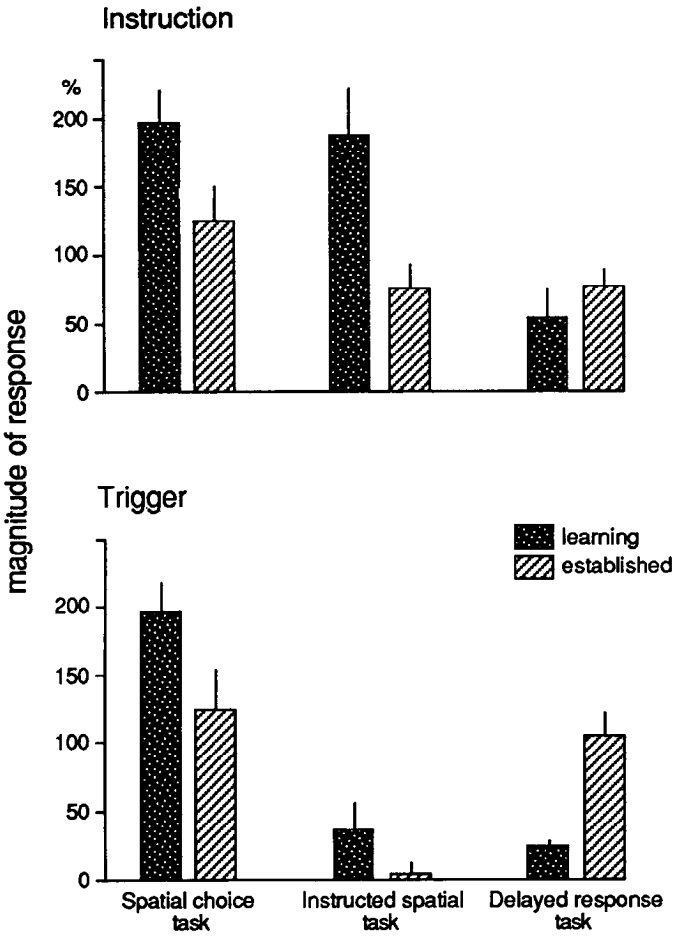


Figure 12. Magnitudes of changes after instruction and trigger stimuli in all neurons tested. The standard time window procedure was employed on all DA neurons tested in both monkeys during learning and established performance of each task, respectively. Differences in magnitudes of changes between learning and established performance were significant for instruction responses in the instructed spatial task ($p < 0.001$, t test), whereas all other comparisons failed to reach this level. See Table 3 for numbers of neurons tested.

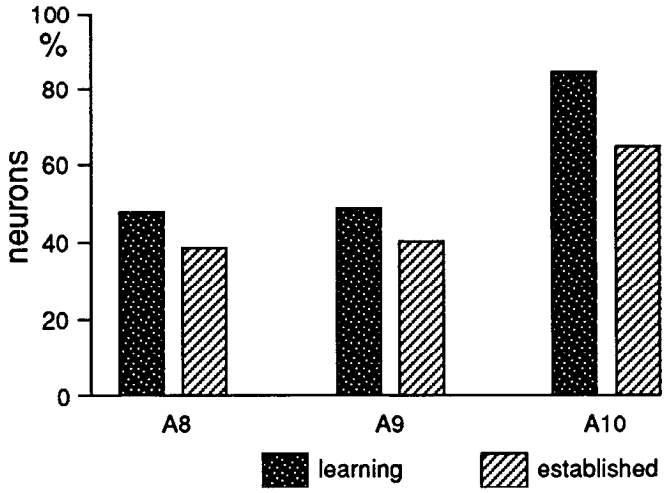


Figure 14. Distribution of DA neurons responding to the instruction cue over midbrain cell groups A8–A10 in both monkeys. Neurons from all three tasks are included. *Ordinate* indicates percentage of responding neurons in each cell group. The fraction of neurons responding during learning is significantly higher in group A10, as compared to groups A8 and A9 ($p < 0.0001$, χ^2 test). See Table 2 for numbers of neurons tested.

the following three tasks in which the arrival of reward and its meaning for task learning are already somewhat familiar.

The fraction of neurons presently responding to reward was also lower than that observed during learning of a delayed alternation task (Ljungberg et al., 1991). This may be due to the double role of reward in that task, serving as both reinforcer and indirect indicator of the next movement target. In agreement with this reasoning, reward responses persisted during established performance of the delayed alternation task, even after overtraining, whereas the present reward responses were largely restricted to the learning phase.

Several arguments suggest that the observed reward responses did not constitute startle responses to stimuli occurring unexpectedly to the animal. Each animal was very familiar with a drop of liquid being delivered at the spout in front of its mouth

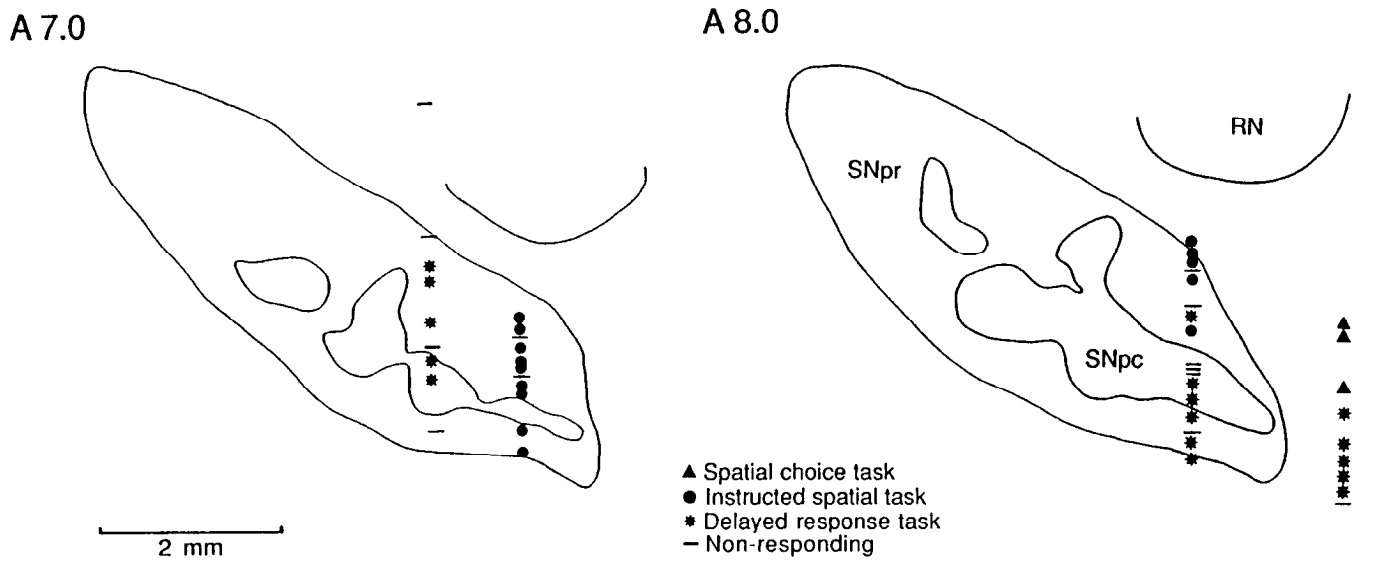


Figure 13. Recording positions of DA neurons with instruction responses during established task performance in one monkey. Positions on two coronal levels are marked separately for each task, whereas nonresponding neurons are commonly labeled by a *dash*. *SNpr* and *SNpc*, pars reticulata and pars compacta of substantia nigra; *RN*, red nucleus.

through its prior learning, performance, and overtraining in the reaction time task with several tens of thousands of trials. Reward responses were largely absent during established task performance and regularly came back with similar frequencies when reward contingencies were changed when advancing from one task to the next one. None of the DA neurons showed response habituation over several tens of consecutive trials, a result that was also seen with all other stimuli effective for activating DA neurons in these tasks. The relation to the delivery of liquid and the lack of relation to stimuli associated with this event are suggested by the effects of closing the liquid tube, as tested in the delayed alternation task (Ljungberg et al., 1991). This manipulation left the noise of the solenoid liquid valve present while blocking liquid flow. DA neurons responding to reward failed to be activated in this situation. A similar specificity for reward was seen before when DA neurons responded to the touch of food reward but failed to be activated by inedible objects (Romo and Schultz, 1990).

An interesting finding was the reproducible depression of activity in error trials occurring in all DA neurons at the precisely the moment when reward would have been delivered in correct trials. The depression did not occur in close temporal relation to a stimulus or a particular behavioral event but appeared as a sharp decrease in activity more than 500 msec after the wrong lever was touched. Reward was normally given at this fixed delay of 500 msec after lever touch throughout the experiment with each animal. It had already been employed for several tens of thousands of trials during learning, established performance, and overtraining of the preceding reaction time task. This suggests that the depression was not a response to the preceding event but reflects a state of expectation of reward at the time of its usual delivery. This would imply that the input to DA neurons had access to centrally stored information concerning the expected time of reward delivery. A similar depression of activity in DA neurons was seen in error trials and in trials with a closed liquid tube during a delayed alternation task (Ljungberg et al., 1991) and when food morsels were replaced by hidden inedible objects (Romo and Schultz, 1990). The visual inspection of rasters and histograms revealed that depressions occurred in many neurons even in the presence of reward responses after correct performance, such that the activating response was added on top of the depressed activity. These depressions were lost together with the activating reward responses after established task performance. This suggests the convergence of two different inputs to DA neurons, one activating input directly driven by the delivery of reward, and one depressant input carrying information about the expected time of arrival of reward but without indicating whether reward was actually delivered.

In conclusion, the restricted occurrence of reward responses during the learning phase and not during established task performance suggests a relationship to the acquisition of new task contingencies. The low responsiveness during established task performance rules out a relationship to general task reinforcement. The fact that many of the same DA neurons also respond to conditioned stimuli important for task performance suggests that reward belongs to a larger class of stimuli effective for activating DA neurons. These stimuli apparently have a particularly alerting and attention-grabbing function for the animal. The particular motivational value of the reception of reward during learning makes this signal a salient stimulus that is very important for the animal when trying to find the new task contingencies. This characteristic appears to determine is capacity to drive DA neurons. Thus, stimuli of prime motivational significance form an important albeit not exclusive class of salient stimuli driving DA neurons. The activity of DA neurons would not selectively transmit information about the reception of primary reward but indicates that a particularly important stimulus occurred in the environment of the subject.

Responses to conditioned stimuli

Whereas reward responses were largely restricted to the learning phase and disappeared with established task performance, responses to the instruction and trigger stimuli occurred throughout both learning and established performance, albeit at varying frequencies. This is probably due to the fact that their behavioral significance was partly known to the animals from the preceding reaction time task. DA neurons failed to respond to known stimuli without any specific meaning for the animal's behavior and only began to respond when the significance of the stimulus became known through learning of the first task (Ljungberg et al., 1992). The change of behavioral significance of familiar conditioned stimuli induced by the modified reward contingencies during the different steps of learning obviously has another impact on DA neurons than the first association of behavioral significance to a familiar but neutral stimulus. The variations of neuronal responsiveness between learning and established performance in each task and between the three tasks should then be related to the changes in behavioral significance of the stimuli during these periods.

Spatial choice task. Animals were familiar with the trigger stimulus from the previous reaction time task (Ljungberg et al., 1992). In the present task, an additional instruction cue was illuminated simultaneously with the trigger stimulus to indicate which of the two levers had to be touched. This visually guided reaching to different targets differed from the single target reaching of the preceding reaction time task and introduced a spatial choice as a first step toward learning the spatial delayed response task. The animal in addition needed to choose between two possible targets in the spatial choice task and thus pay more attention to the conditioned stimuli. Whereas neuronal responses to the trigger stimulus were very low after overtraining of the preceding reaction time task, they reappeared when a spatial choice was required in the present tasks. The increased responsiveness should mainly be due to the increased alerting and attention-grabbing function of the reward-predicting trigger stimulus induced by the changed behavioral situation. We have suggested that responses to triggering stimuli were due to the reward-predicting nature of this conditioned stimulus (Schultz and Romo, 1990; Ljungberg et al., 1992). The present responses to the combined instruction–trigger stimulus are probably not due to the higher sensory impact of two simultaneous visual stimuli, because responses declined again when task performance was established. They should also not be related to the specific spatial information contained in the instruction because they lacked spatial specificity.

Instructed spatial task. As the second step toward learning the spatial delayed response task, the instruction cue indicating the target of reaching was separated by a fixed period of 1 sec from the following trigger stimulus, which determined the moment of reaching. The animal learned to withhold reaching after the instruction cue and wait until the trigger stimulus occurred. This change in the behavioral situation had two effects on DA neurons. Neuronal responsiveness increased again during learn-

ing of this task, and the responses occurred virtually exclusively to the instruction cue and not to the trigger stimulus. Animals reacted with a saccadic eye movement toward the position of the instruction-lever compound and fixated this target until the reaching movement occurred, whereas the trigger stimulus failed to elicit an eye movement. Also, the arm movement began in several trials too soon after the trigger to be elicited by this stimulus. This suggests that the instruction, besides indicating the spatial position, served to trigger a delayed arm movement and that the trigger light had largely lost its unique triggering function and merely served as a temporal reference for the delayed movement. Thus, the neuronal responses to the instruction cue and their absence after the trigger stimulus should be explained by a transfer of the alerting, attention-grabbing, and reward-predicting characteristics from the trigger stimulus to the instruction cue.

Spatial delayed response task. The last step in learning this task involved two modifications. The instruction–trigger interval was randomized and prolonged to 2.5–3.5 sec. A delay was introduced between instruction offset and trigger onset during which task specific information was absent and the animal had to memorize the spatial position of the target lever indicated by the instruction. The response of DA neurons became immediately separated into a response to the instruction and, in the same trial, a response to the trigger stimulus with very similar temporal characteristics. This resulted in a lowering of responsiveness to each stimulus. Responses increased with established task performance, and a slightly higher fraction of neurons responded to the trigger as compared to the instruction cue. The study of eye and arm movements indicated that the two signals had attained different characteristics. The instruction cue was the first stimulus in each trial and also provided important spatial information for correct task performance. It elicited a saccadic eye movement. The trigger stimulus was again a distinctive stimulus that elicited a separate eye movement and the reaching movement toward the target. As principal reward predictor, the trigger stimulus was the main incentive stimulus of the task. However, the instruction cue through its close temporal association with the trigger stimulus and reward may not have been completely devoid of incentive properties.

It is presently unclear whether the stimulus properties of the instruction cue are due to its general role in the delayed response task or are derived from the particular training procedure employed. It also remains to be shown to what extent instruction responses would be reduced with overtraining. In a previous experiment, instruction responses occurred in an asymmetrically reinforced delayed go–no-go task only when the task was occasionally employed, whereas they disappeared with frequent use (Schultz and Romo, 1990). A systematic investigation of responses to instructions in different behavioral contexts would be an interesting way to study the relationship of DA neurons to stimuli not associated with the immediate obtainment of reward.

Role of DA neurons in delay tasks

Previous studies investigating the role of DA in animal cognition found that local DA depletions or local administration of DA receptor antagonists into the prefrontal cortex of monkeys result in impaired performance of spatial delayed response tasks (Brozoski et al., 1979; Sawaguchi and Goldman-Rakic, 1991). During the performance of delayed response tasks similar to the one used presently, prefrontal neurons respond phasically to spatial instruction cues that are being memorized, and to trigger stimuli that induce the recall of stored information and elicit the behavioral response (Fuster, 1973; Kubota et al., 1974; Niki, 1974). Prefrontal neurons also show sustained activity during the instruction–trigger delay, which may reflect processing related to representational task components, such as working memory, expectation of the trigger stimulus, and preparation of the behavioral response. Similar activity is found in striatal neurons during the performance of delay tasks (Alexander, 1987; Schultz and Romo, 1988), including the same spatial delayed response task in the two monkeys used presently (Apicella et al., 1992). DA receptor antagonists reduce and DA itself increases both the phasic responses and the sustained delay activity of prefrontal neurons when applied iontophoretically (Sawaguchi et al., 1990a,b). These combined findings would explain the deficits in delayed response performance after impaired DA transmission and suggest that prefrontal neuronal activity is directly responsible for the performance of delayed response tasks. By contrast, DA neurons failed to show sustained activations during the delay, suggesting that they do not encode the abovementioned representational task components. The phasic responses to instruction and trigger stimuli were not specific for the spatial component of task performance but were apparently related to the salience of these stimuli. Thus, DA neurons appear to enable prefrontal neuronal activity related to representational task components underlying delayed response performance without specifically encoding these components. The mechanisms of this enabling function are presently unknown. It is possible that sustained neuronal activity is facilitated by the action of ionic channels with slow time constants in cortical and striatal neurons (Stafstrom et al., 1982; Calabresi et al., 1987; Kawaguchi et al., 1989; Pennartz et al., 1991) and is built up by cortico–basal ganglia–cortical loops (Romo and Schultz, in press). It would be very interesting to see whether DA may have a specific influence on the time constants of these channels and thus have a regulatory effect on the passage of neuronal activity in these loops.

References

Alexander GE (1987) Selective neuronal discharge in monkey putamen reflects intended direction of planned limb movements. Exp Brain Res 67:623–634.

Apicella P, Scarnati E, Ljungberg T, Schultz W (1992) Neuronal activity in monkey striatum related to the expectation of predictable environmental events. J Neurophysiol 68:945–960.

Brozoski TJ, Brown RM, Rosvold HE, Goldman PS (1979) Cognitive deficit caused by regional depletion of dopamine in prefrontal cortex of rhesus monkey. Science 205:929–932.

Calabresi P, Mercuri N, Stanzione P, Stefani A, Bernardi G (1987) Intracellular studies on the dopamine-induced firing inhibition of neostriatal neurons *in vitro*: evidence for D1 receptor involvement. Neuroscience 20:757–771.

Cools AR, Van den Bercken JHL, Horstink WI, Van Spaendonck KPM, Berger HJC (1984) Cognitive and motor shifting aptitude disorder in Parkinson's disease. J Neurol Neurosurg Psychiatry 47:443–453.

Felten DL, Sladeck JR (1983) Monoamine distribution in primate brain. V. Monoaminergic nuclei: anatomy, pathways and local organisation. Brain Res Bull 10:171–284.

Fuster JM (1973) Unit activity of prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. J Neurophysiol 36:61–78.

Hellweg FC, Schultz W, Creutzfeldt OD (1977) Extracellular and intracellular recordings from cat's cortical whisker projection area: thalamocortical response transformation. J Neurophysiol 40:462–479.

Kawaguchi Y, Wilson CJ, Emson PC (1989) Intracellular recording of identified neostriatal patch and matrix spiny cells in a slice preparation preserving cortical inputs. J Neurophysiol 62:1052–1068.

Kubota K, Iwamoto T, Suzuki H (1974) Visuokinetic activities of primate prefrontal neurons during delayed-response performance. J Neurophysiol 37:1197–1212.

Ljungberg T, Apicella P, Schultz W (1991) Responses of monkey midbrain dopamine neurons during delayed alternation performance. Brain Res 586:337–341.

Ljungberg T, Apicella P, Schultz W (1992) Responses of monkey dopamine neurons during learning of behavioral reactions. J Neurophysiol 67:145–163.

Miller JD, Sanghera MK, German DC (1981) Mesencephalic dopaminergic unit activity in the behaviorally conditioned rat. Life Sci 29:1255–1263.

Niki H (1974) Differential activity of prefrontal units during right and left delayed response trials. Brain Res 70:346–349.

Pennartz CMA, Boeijinga PH, Kitai ST, Lopes da Silva FH (1991) Contribution of NMDA receptors to postsynaptic potentials and paired-pulse facilitation in identified neurons of the rat nucleus accumbens *in vitro*. Exp Brain Res 86:190–198.

Porrino LJ, Goldman-Rakic PS (1982) Brainstem innervation of prefrontal and anterior cingulate cortex in the rhesus monkey revealed by retrograde transport of HRP. J Comp Neurol 205:63–76.

Romo R, Schultz W (1990) Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. J Neurophysiol 63:592–606.

Romo R, Schultz W (in press) Role of primate basal ganglia and frontal cortex in the internal generation of movements: neuronal activity in the supplementary motor area. Exp Brain Res, in press.

Saint-Cyr JA, Taylor AE, Lang AE (1988) Procedural learning and neostriatal function in man. Brain 111:941–959.

Sawaguchi T, Goldman-Rakic PS (1991) D1 Dopamine receptors in prefrontal cortex: involvement in working memory. Science 251:947–950.

Sawaguchi T, Matsumura M, Kubota K (1990a) Catecholaminergic effects on neuronal activity related to a delayed response task in monkey prefrontal cortex. J Neurophysiol 63:1385–1400.

Sawaguchi T, Matsumura M, Kubota K (1990b) Effects of dopamine antagonists on neuronal activity related to a delayed response task in monkey prefrontal cortex. J Neurophysiol 63:1401–1412.

Schneider JS, Kovelowski CJ (1990) Chronic exposure to low doses of MPTP. I. Cognitive deficits in motor asymptomatic monkeys. Brain Res 519:122–128.

Schultz W (1986) Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. J Neurophysiol 56:1439–1462.

Schultz W, Romo R (1987) Responses of nigrostriatal dopamine neurons to high intensity somatosensory stimulation in the anesthetized monkey. J Neurophysiol 57:201–217.

Schultz W, Romo R (1988) Neuronal activity in the monkey striatum during the initiation of movements. Exp Brain Res 71:431–436.

Schultz W, Romo R (1990) Dopamine neurons of the monkey midbrain: contingencies of responses to stimuli eliciting immediate behavioral reactions. J Neurophysiol 63:607–624.

Schultz W, Ljungberg T, Apicella P (1991) Activity of dopamine neurons in monkeys learning and performing cognitive tasks. Soc Neurosci Abstr 17:1218.

Simon H, Scatton B, LeMoal M (1980) Dopaminergic A10 neurons are involved in cognitive functions. Nature 286:150–151.

Stafstrom CE, Schwindt PC, Crill WE (1982) Negative slope conductance due to a persistent subthreshold sodium current in cat neocortical neurons *in vitro*. Brain Res 236:221–226.

Steinfels GF, Heym J, Strecker RE, Jacobs BL (1983) Behavioral correlates of dopaminergic unit activity in freely moving cats. Brain Res 258:217–228.

Szabo J (1980) Organization of the ascending striatal afferents in monkeys. J Comp Neurol 189:307–321.

Behavioral/Systems/Cognitive

Earlier Development of the Accumbens Relative to Orbitofrontal Cortex Might Underlie Risk-Taking Behavior in Adolescents

Adriana Galvan,[1] Todd A. Hare,[1] Cindy E. Parra,[1] Jackie Penn,[1] Henning Voss,[1] Gary Glover,[2] and B. J. Casey[1]

[1]The Sackler Institute for Developmental Psychobiology, Weill Medical College of Cornell University, New York, New York 10021, and [2]Department of Radiology and Neurosciences Program, Center for Advanced Magnetic Resonance Technology at Stanford University, Stanford, California 94305

Adolescence has been characterized by risk-taking behaviors that can lead to fatal outcomes. This study examined the neurobiological development of neural systems implicated in reward-seeking behaviors. Thirty-seven participants (7–29 years of age) were scanned using event-related functional magnetic resonance imaging and a paradigm that parametrically manipulated reward values. The results show exaggerated accumbens activity, relative to prefrontal activity in adolescents, compared with children and adults, which appeared to be driven by different time courses of development for these regions. Accumbens activity in adolescents looked like that of adults in both extent of activity and sensitivity to reward values, although the magnitude of activity was exaggerated. In contrast, the extent of orbital frontal cortex activity in adolescents looked more like that of children than adults, with less focal patterns of activity. These findings suggest that maturing subcortical systems become disproportionately activated relative to later maturing top–down control systems, biasing the adolescent's action toward immediate over long-term gains.

Key words: **accumbens; adolescent; reward; development; orbital frontal cortex; risk-taking**

Introduction

Onset of substance abuse often occurs during the increased risk-taking period of adolescence (Silveri et al., 2004). Little is known to date about neurobiological factors that may predispose adolescents to increased risk-taking behaviors. In adults, mesolimbic regions have been implicated in reward (Knutson et al., 2001; Elliott et al., 2003; McClure et al., 2004), risk-taking (Kuhnen and Knutson, 2005), and addiction (Hyman and Malenka, 2001; Volkow et al., 2004), but less is known about the development of these systems. The purpose of this study was to test the hypothesis that adolescence is a developmental period of increased responsivity to reward relative to childhood and adulthood. Specifically, we examined whether differences in the development of subcortical [e.g., nucleus accumbens (NAcc)] relative to prefrontal [e.g., orbital frontal cortex (OFC)] regions may characterize this period of development to help explain the increase in risk-taking behavior.

Adolescence is characterized by continued structural and functional development of frontostriatal circuitry implicated in behavioral regulation. Periadolescent rats show increases in reward-related dopamine transmission in the striatum (Laviola et al., 1999), and nonhuman primates show increased dopaminergic innervation in the prefrontal cortex (PFC) (Rosenberg and Lewis, 1994, 1995). Human imaging studies show frontostriatal region changes (Giedd et al., 1999; Sowell et al., 1999; Casey et al., 2005) that seem to parallel increased cognitive control (Casey et al., 1997; Rubia et al., 2000; Luna et al., 2001; Luna and Sweeney, 2004; Steinberg, 2004). These changes appear to show activation shifts of prefrontal regions from diffuse to more focal recruitment over time (Casey et al., 1997; Bunge et al., 2002; Moses et al., 2002; Durston et al., 2006). Neuroimaging studies cannot definitively characterize the mechanism of such developmental changes (e.g., synaptic pruning, myelination). However, these volume and structural changes may reflect refinement and fine-tuning of reciprocal projections from these brain regions during maturation. Thus, this interpretation is only speculative.

Recently, neuroimaging studies have begun to examine reward-related processing in adolescents and have shown NAcc activation as shown in adults (Bjork et al., 2004; May et al., 2004; Ernst et al., 2005). However, the results have been mixed as to how adolescents and adults differ in activity. These studies have focused primarily on the region of the accumbens rather than OFC in examining changes. Furthermore, little attention has been given to characterizing NAcc and OFC development from childhood through adulthood. Tracking this development provides additional constraints on whether changes reported in adolescence are specific to this period of development or reflect maturational linear changes.

Here, we used functional magnetic resonance imaging (fMRI) to examine behavioral and neural responses to reward value ma-

Received Jan. 5, 2006; revised May 15, 2006; accepted May 25, 2006.

This work was supported in part by National Institute on Drug Abuse Grants R01 DA18879 and R21 DA15882, National Institute of Mental Health Grant P50 MH62196 (B.J.C.), and National Eye Institute Grant T32 EY07138 fellowship (A.G.). We gratefully acknowledge all of the participants and their families for participating in this study and three anonymous reviewers.

Correspondence should be addressed to either Adriana Galvan or B. J. Casey, 1300 York Avenue, Box 140, New York, NY 10021. E-mail: adg2006@med.cornell.edu or bjc2002@med.cornell.edu.

DOI:10.1523/JNEUROSCI.1062-06.2006

Copyright © 2006 Society for Neuroscience 0270-6474/06/266885-08$15.00/0

nipulations across development. We focused on the NAcc and OFC given previous reports in animal (Hikosaka and Watanabe, 2000; Pecina et al., 2003), imaging (O'Doherty et al., 2001; Zald et al., 2004), and addiction (Hyman and Malenka, 2001) studies implicating them in reward-related learning. Based on rodent models (Laviola et al., 1999; Spear, 2000) and previous imaging work (Ernst et al., 2005), we hypothesized that, relative to children and adults, adolescents would show an exaggerated accumbens response suggestive of refined focal activations within the accumbens during this period, in concert with less mature activations in top–down PFC regions.

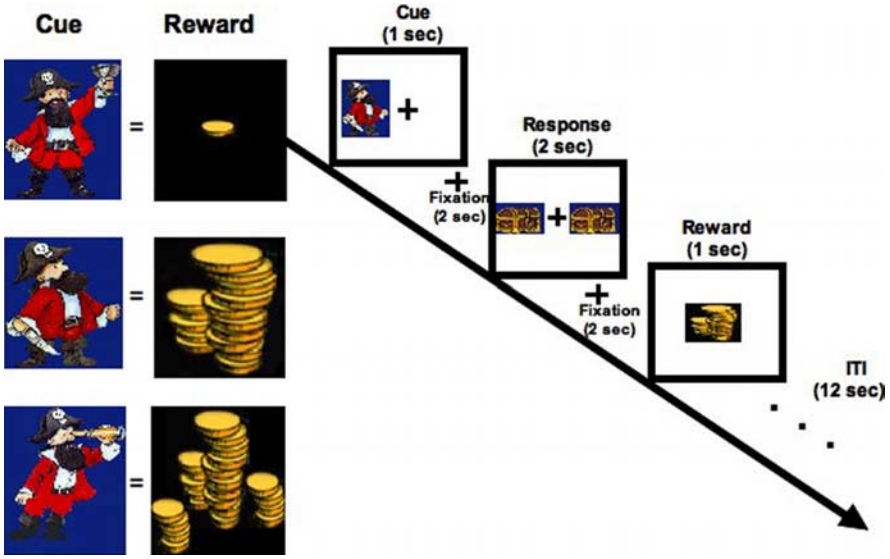


Figure 1. Behavioral paradigm. Left panel, Three cues were each paired with a distinct reward value (counterbalanced across subjects) that remained constant throughout the experiment. Right panel, The paradigm consisted of a cue, response, and reward that were temporally separated in time with 12 s ITI. Total trial length was 20 s.

Materials and Methods

Participants. Sixteen children (seven females; aged 7–11, with mean age of 9.8 years), 13 adolescents (six females; aged 13–17, with mean age of 16 years), and 12 right-handed healthy adults (six females; aged 23–29, with mean age of 25 years) participated in the fMRI experiment. A separate statistical analysis on the adult data was reported previously (Galvan et al., 2005). Three children and one adolescent were excluded from the analysis due to excessive motion (>2 mm). Motion was >0.5 voxels (1.56 mm) in any direction for two subjects (one child and one adult) included in the analysis. Eliminating these subjects from the analysis did not change the results and the different age groups did not differ significantly in in-plane motion (adults: $x = 0.48$, $y = 0.76$, $z = 0.49$; adolescents: $x = 0.26$, $y = 0.58$, $z = 0.45$; children: $x = 0.18$, $y = 0.76$, $z = 0.36$). Subjects had no history of neurological or psychiatric disorder and each gave informed consent (parental consent and child assent for adolescents and children) for a protocol approved by the Institutional Review Board of Weill Cornell Medical College of Cornell University. The experiment on adolescents and children was simulated in a mock scanner before the experiment, in which they were exposed to the sounds they would hear during the actual experiment.

Experimental task. Participants were tested using an adapted version of a delayed response two-choice task previously used in nonhuman primates (Cromwell and Schultz, 2003) and described previously (Galvan et al., 2005) in an event-related fMRI study (Fig. 1). In this task, three cues (counterbalanced) were each associated with a distinct reward value. Subjects were instructed to press either their index or middle finger to indicate the side on which a cue appeared when prompted, and to respond as quickly as possible without making mistakes.

The stimulus parameters were as follows. One of three pirate cartoon images was presented in pseudorandom order on either the left or right side of a centered fixation for 1000 ms (Fig. 1). After a 2000 ms delay, subjects were presented with a response prompt of two treasure chests on both sides of the fixation (2000 ms) and instructed to press a button with their right index finger if the pirate was on the left side of the fixation or their right middle finger if the pirate was on the right side of the fixation. After another 2000 ms delay, reward feedback of either a small, medium, or large amount of coins was presented in the center of the screen (1000 ms). Each pirate was associated with a distinct reward amount. There was a 12 s intertrial interval (ITI) before the start of the next trial. Total trial length was 20 s. Subjects were not rewarded if they failed to make a response or if they made an error; in both cases, they received an error message at the time they would normally receive reward feedback.

Subjects were guaranteed $50 for participation in the study and were told they could earn up to $25 more, depending on performance (as indexed by reaction time and accuracy) on the task. Although the reward amounts were distinctly different from one another, the exact value of each reward was not disclosed to the subject, because during pilot studies, subjects reported counting the money after each trial and we wanted to avoid this possible distraction. Stimuli were presented with the integrated functional imaging system (PST, Pittsburgh, PA) using a liquid crystal display video display in the bore of the magnetic resonance (MR) scanner and a fiber optic response collection device.

The experiment consisted of five runs of 18 trials (six each of small, medium, and large reward trials), which lasted 6 min and 8 s each. Each run had six trials of each reward value presented in random order. At the end of each run, subjects were updated on how much money they had earned during that run. The amount of money earned was consistent across all subjects and all received a continuous schedule of reinforcement (rewarded on 100% of trials). Before beginning the experiment, subjects were shown the actual money they could earn to ensure motivation. They received detailed instructions that included familiarization with the stimuli used. For instance, subjects were shown the three cues and three reward amounts they would be seeing during the experiment. They were not told how the cues related to the rewards. We explicitly emphasized that there were three amounts of reward, one being small, another medium, and another large. These amounts are visually obvious in the experiment because the number of coins in the stimuli increases with increasing reward. Only one subject could articulate the association between specific stimuli and reward amounts, when asked explicitly about this association during debriefing of the subject at the end of the experiment.

Image acquisition. Imaging was performed using a 3T General Electric (Milwaukee, WI) MRI scanner using a quadrature head coil. Functional scans were acquired using a spiral in and out sequence (Glover and Thomason, 2004). The parameters included the following: repetition time (TR), 2000 ms; echo time (TE), 30 ms; 64 × 64 matrix; 29 5 mm coronal slices; 3.125 × 3.125 mm in-plane resolution; flip, 90° for 184 repetitions, including four discarded acquisitions at the beginning of each run. Anatomical T1-weighted in-plane scans were collected (TR, 500; TE, min; 256 × 256; field of view, 200 mm; 5 mm slice thickness) in the same locations as the functional images in addition to a three-dimensional data set of high-resolution spoiled gradient-recalled acquisition in a steady state images (TR, 25; TE, 5; 1.5 mm slice thickness; 124 slices).

Image analysis. The Brainvoyager QX (Brain Innovations, Maastricht, The Netherlands) software package was used to perform a random-effects analysis of the imaging data. Before analysis, the following preprocessing procedures were performed on the raw images: three-dimensional motion correction to detect and correct for small head movements by spatial alignment of all volumes to the first volume by

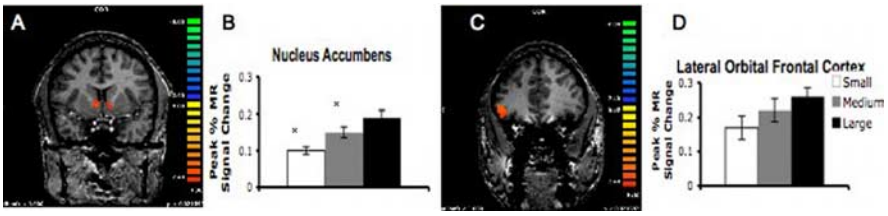


Figure 2. Localization of nucleus accumbens (***A***) and orbital frontal cortex (***C***) activation to reward. There was a main effect of reward value in the nucleus accumbens (***B***) [right ($x = 6, y = 5, z = -2$) and left ($x = -8, y = 6, z = -2$)] but not in the right lateral orbital frontal cortex ($x = 46, y = 31, z = 1$) (***D***). Error bars indicate SEM. Asterisks denote significant differences between small and medium, medium and large, and small and large.

rigid body transformation, slice scan time correction (using sinc interpolation), linear trend removal, high-pass temporal filtering to remove nonlinear drifts of three or fewer cycles per time course, and spatial data smoothing using a Gaussian kernel with a 4 mm full width at half-maximum. Estimated rotation and translation movements never exceeded 2 mm for subjects included in this analysis. Functional data were coregistered to the anatomical volume by alignment of corresponding points and manual adjustments to obtain optimal fit by visual inspection and were then transformed into Talairach space. Functional voxels were interpolated from the acquisition voxel size of 48.83 mm^3 to a resolution of 1 mm^3 during Talairach transformation. The NAcc and OFC were defined by Talairach coordinates in conjunction with reference to the Duvernoy brain atlas (Talairach and Tournoux, 1988; Duvernoy, 1999).

The initial omnibus general linear model (GLM) analysis included all subjects and all runs across the entire trial (relative to pretrial baseline) to determine regions sensitive to reward (NAcc and OFC). To ensure that statistical analyses were performed on the same regions for each age group, separate GLM analyses were performed. Each group showed activation in the NAcc and OFC based on a reward versus baseline contrast. Localization of these regions was further confirmed for each group separately by Talairach coordinates in conjunction with reference to the Duvernoy brain atlas (Talairach and Tournoux, 1988; Duvernoy, 1999) as described above. Previous methodological work has shown that the stereotactic registration and time course of the hemodynamic response across the ages tested in the current study are not dissimilar (Burgund et al., 2002; Kang et al., 2003). Subsequent analysis and *post hoc* contrasts were performed on the regions identified with this initial omnibus GLM for all groups together and then separately for each group. Last, a conjunction analysis was performed that identified the voxels that were commonly activated across all three groups, in the NAcc and OFC (supplemental Fig. 1, available at www.jneurosci.org as supplemental material). The regions of interest identified in the conjunction analysis overlapped with those identified with the initial omnibus GLM, and *post hoc* tests confirmed similar effects as those obtained with the above analyses.

In the whole-group analysis, the omnibus GLM was comprised of all runs across the entire trial (5 runs × 37 subjects = 185 *z*-normalized functional time courses) and was conducted with reward magnitude as the primary predictor. The predictors were obtained by convolution of an ideal boxcar response (assuming a value of 1 for the volume of task presentation and a volume of 0 for the remaining time points) with a linear model of the hemodynamic response (Boynton et al., 1996) and used to build the design matrix of each time course in the experiment. Only correct trials were included and separate predictors were created for error trials. The total number of correct trials for each group was as follows: 1130 for children ($n = 13$), 1061 for adolescents ($n = 12$), and 1067 for adults ($n = 12$). The fewer number of trials for children was corrected by including an additional child subject.

Post hoc contrast analyses were then performed based on *t* tests on the β weights of predictors to identify a region of interest in the NAcc and OFC. Contrasts were conducted with a random-effects analysis. Time series and percent changes in MR signal, across each data point of the entire trial (18 s) relative to 2 s of fixation preceding trial onset (total trial duration was 20 s), were calculated using event-related averaging over significantly active voxels obtained from the contrast analyses. The calculation of number of voxels recruited in each region by age group was based on the GLM analyses conducted on each group described above.

Corrections for multiple comparisons were based on Monte Carlo simulations, which were run using the AlphaSim program within AFNI (Cox, 1996), to determine appropriate contiguity thresholds to achieve a corrected α level of $p < 0.01$ (Forman et al., 1995) based on a search volume of 450 mm^3 for the NAcc. A corrected α level of $p < 0.05$ in the OFC was based on a search volume of ~25,400 mm^3 (Forman et al., 1995). OFC activation did not survive the more stringent threshold of $p < 0.01$ across groups.[a]

Results

Imaging results

The omnibus GLM analysis of the imaging data identified the NAcc [right ($x = 6, y = 5, z = -2$) and left ($x = -8, y = 6, z = -2$)] and right OFC ($x = 46, y = 31, z = 1$) depicted in Figure 2, *A* and *C*, with reward value as the primary predictor, across all subjects and runs of the experiment for the entire trial (18 s), relative to the 2 s intertrial interval preceding the start of the next trial (e.g., reward vs baseline contrast). Within these regions, there was a main effect of reward value ($F_{(2,72)} = 8.424$; $p = 0.001$) (Fig. 2*B*) in the NAcc, but not in the OFC ($F_{(2,72)} = 1.3$; $p = 0.44$) (Fig. 2*D*). *Post hoc t* tests on the main effect of reward for the NAcc confirmed significant differences between the large and small ($t_{(36)} = 4.35$; $p < 0.001$), large and medium ($t_{(36)} = 2.01$; $p < 0.05$), and medium and small ($t_{(36)} = 2.09$; $p < 0.04$) rewards, with greater activation for greater rewards.

Developmental differences in magnitude and extent of activity to reward

Because the focus of this study was on how reward influences neural recruitment across development, we examined developmental differences in the magnitude and extent of accumbens and OFC activity for trials to the largest reward. The magnitude of activity was calculated as a percent change in MR signal averaged across the first 18 s of the trial relative to the intertrial interval of fixation immediately preceding the trial (2 s), which was averaged across the entire experiment (90 trials = 900 scans). This calculation was performed for each group. The extent of activity was calculated as the volume of activity (number of voxels) across runs, by group, using the same contrast.

Magnitude of activity. In the accumbens and OFC, there were significant developmental differences in percent change in MR signal ($F_{(2,22)} = 6.47$, $p < 0.01$; $F_{(2,22)} = 5.02$, $p = 0.01$, respectively) (Fig. 3*A*,*B*). In the accumbens, adolescents showed the greatest signal change. *Post hoc* tests confirmed significant differences between adolescents and children ($t_{(11)} = 4.2$; $p = 0.03$) and between adolescents and adults ($t_{(11)} = 5.5$; $p = 0.01$) in magnitude of accumbens activity. In the OFC, *post hoc* tests confirmed significant differences between children and adolescents ($t_{(11)} = 4.9$; $p = 0.01$) and children and adults ($t_{(11)} = 3.99$; $p = 0.01$). Thus, adolescents showed enhanced activity in the accumbens and this pattern differed from that in the OFC and from children and adults.

Extent of activity. There were significant developmental differ-

[a]Cluster sizes of 6 and 10 for the accumbens and OFC, respectively, were determined by these simulations. Cluster sizes of 8 and 10 in the adolescent and child data, respectively, survived more stringent thresholds ($p < 0.002$ and $p < 0.001$, respectively). In the OFC, cluster sizes of 14 and 18 in adolescents and children, respectively, survived more stringent thresholds ($p < 0.004$ and $p < 0.001$, respectively).

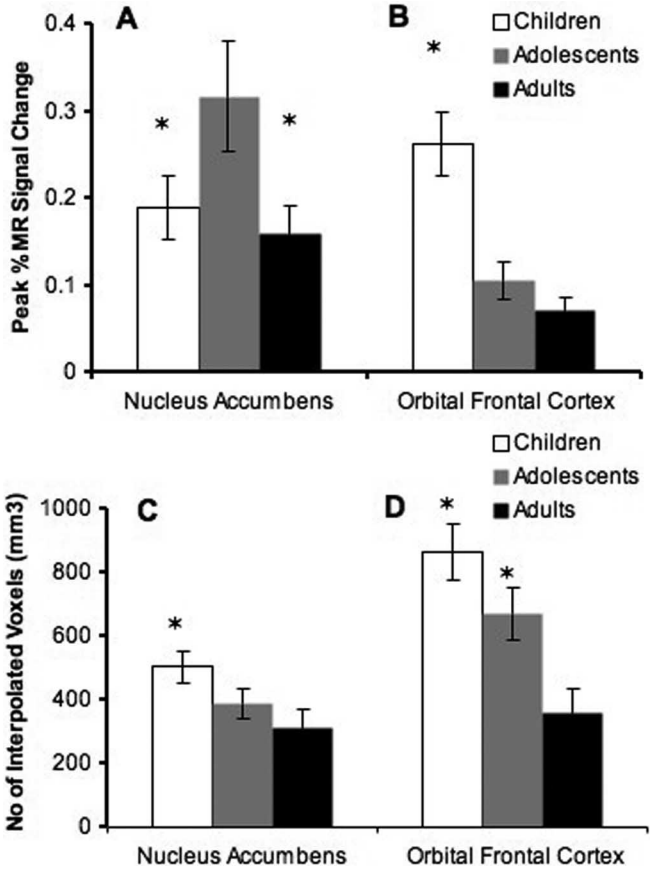


Figure 3. Magnitude and extent of accumbens and OFC activity to reward. ***A***, Adolescents showed exaggerated percent change in MR signal to large reward relative to children and adults in the accumbens. ***B***, In the OFC, children had the greatest percent change in MR signal relative to adolescents and adults. ***C***, Children showed the largest volume of activity in the accumbens relative to adolescents and adults. ***D***, Children and adolescents showed greater volume of activity in the OFC relative to adults. Error bars indicate SEM. Asterisks denote significant activation differences between children and adolescents and adolescents and adults in ***A***; greater activation in children relative to adolescents and adults in ***B***; greater volume of activity in children relative to adolescents and adults in ***C***; and greater volume of activity in children relative to adolescents and adolescents relative to adults in ***D***.

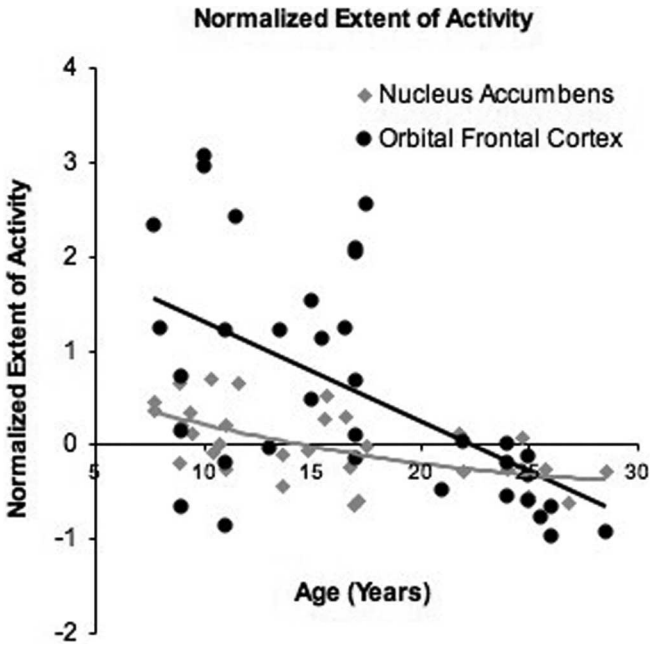


Figure 4. Normalized extent of activity measure for the nucleus accumbens and OFC for all subjects, adjusted for the average extent of activity (x − mean/mean) for each region.

ences in the extent of activity in the accumbens ($F_{(2,22)} = 4.7$; $p < 0.02$) and OFC ($F_{(2,22)} = 5.01$; $p = 0.01$). *Post hoc* tests confirmed the largest volume of activity in the accumbens for children (503 ± 43 interpolated voxels) relative to adolescents (389 ± 71 interpolated voxels) ($t_{(22)} = 4.2$; $p < 0.05$) and adults (311 ± 84 interpolated voxels) ($t_{(22)} = 3.4$; $p < 0.05$) (Fig. 3*C*). Adolescents and adults did not differ ($t_{(22)} = 0.87$; $p = 0.31$). For the OFC, children (864 ± 165 interpolated voxels) ($t_{(22)} = 7.1$; $p = 0.01$) and adolescents (671 ± 54) ($t_{(22)} = 5.8$; $p = 0.01$) showed the largest extent of activity relative to adults (361 ± 45 voxels) (Fig. 3*D*), but there were no significant differences between children and adolescents ($t_{(22)} = 1.8$; $p = 0.07$). This pattern of activity reflects protracted development of OFC relative to the NAcc (Fig. 4, graph).

Developmental differences in temporal processing of reward value

To examine differential changes in neural recruitment throughout the experiment, we examined the main effect of, and interactions with, time (early, middle, and late trials) on MR signal change in the NAcc or OFC. The effect of time was observed only in the interaction of time by group by reward in the accumbens ($F_{(8,136)} = 3.08$; $p = 0.003$) and less robustly in the OFC ($F_{(8,136)} = 2.71$; $p = 0.02$). This interaction was driven primarily by changes occurring during the late trials of the experiment (for changes as a function of early, middle, and late trials, see supplemental Fig. 2, available at www.jneurosci.org as supplemental material). Figures 5 and 6 depict the temporal course of change in MR signal as a function of small, medium, and large reward values for late trials by group for each region. These time series show an exaggerated change in accumbens activity in adolescents relative to children or adults for the small and large reward trials that occurs ~5–6 s after the response and the point in which all three age groups show a change in MR signal. This pattern is illustrated graphically in Figure 7 for clarity (for change in OFC activity at this time point for all three age groups, see supplemental Fig. 3, available at www.jneurosci.org as supplemental material).

Behavioral results

The effects of time on task and reward value were tested with a 5 (runs) × 3 (small, medium and large reward) × 3 (group) ANOVA for the dependent variables of mean reaction time for correct trials and mean accuracy. There were main effects of reward value ($F_{(2,72)} = 9.51$; $p = 0.001$) and group ($F_{(2,220)} = 4.37$; $p = 0.02$) and significant interactions of reward by time ($F_{(8,288)} = 4.176$; $p < 0.001$) and group by reward by time ($F_{(16,272)} = 3.01$; $p = 0.01$) for mean reaction time. The main effect of reward showed that, across all subjects, mean reaction times were faster to the largest reward (mean, 515.47; SD, 178.75; $t_{(36)} = 3.8$; $p < 0.001$) relative to medium (mean, 556.89; SD, 180.53) or small reward (mean, 552.39; SD, 180.35). The significant interaction of reward by time was driven primarily by the three-way interaction of group by reward by time. Adults differed in mean reaction time to all three reward values by the end of the experiment (Fig. 8). Adolescents were significantly faster to the large relative to the medium and small rewards with no difference between the medium and small rewards. The children showed no significant differences in mean reaction time to the small, medium, or large rewards. There were no significant correlations

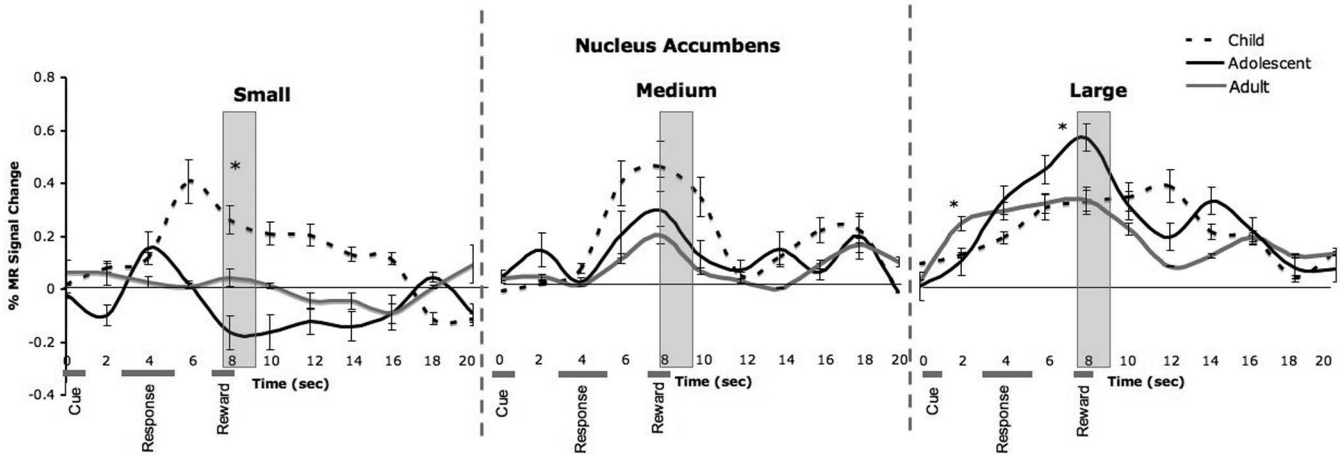


Figure 5. Temporal changes in the nucleus accumbens as a function of small, medium, and large reward values for late trials of the experiment for each age group. The gray bars correspond to the point(s) ~5–6 s after the response. Error bars indicate SEM. Asterisks denote activation differences between groups.

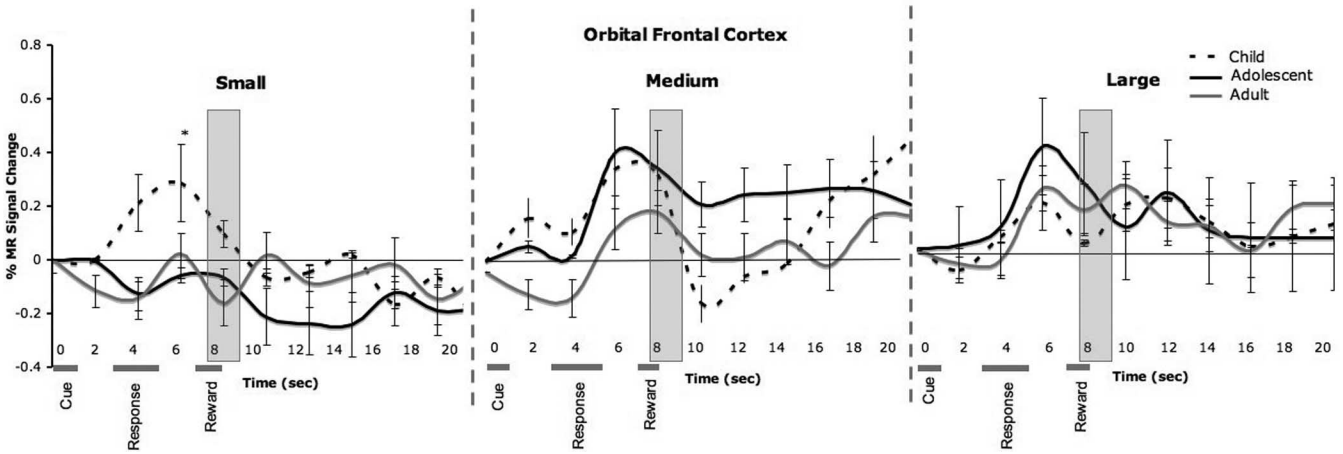


Figure 6. Temporal changes in the orbitofrontal cortex as a function of small, medium and large reward values for late trials of the experiment for each age group. The gray bars correspond to the point(s) ~5–6 s after the response. Error bars indicate SEM. Asterisks denote activation differences between groups.

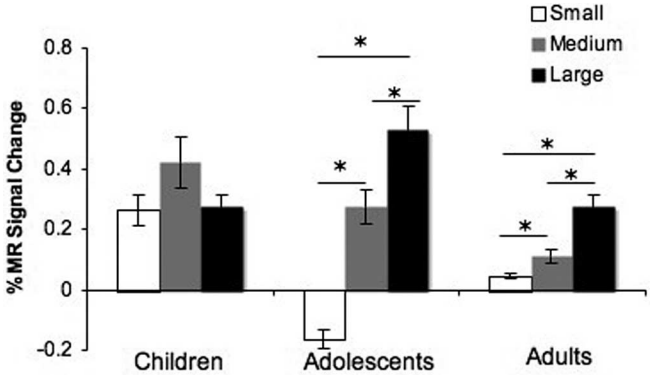


Figure 7. Percent change in MR signal ~5–6 s after the response relative to pretrial baseline for each age group, showing an exaggerated change in accumbens activity in adolescents relative to children or adults for the small and large reward trials. Error bars indicate SEM. Asterisks denote activation differences between groups.

between mean reaction time or accuracy and accumbens or orbitofrontal activity.

There were no significant effects of reward ($F_{(2,72)} = 0.26$; $p = 0.40$), group ($F_{(2,220)} = 0.73$; $p = 0.80$), or time ($F_{(4,476)} = 0.57$; $p = 0.44$) or interactions for mean accuracy. All subjects had high accuracy across reward values (children: small, 96%; medium, 98%; large 96%; adolescents: small, 98%; medium, 99%; large, 99%; and adults: small, 98%; medium, 99%; large, 99%).

Discussion

This study examined behavioral and neural responses to reward value manipulations across development. Our findings support our hypothesis that adolescents differ from children and adults in NAcc and OFC recruitment, regions previously implicated in reward processing (Knutson et al., 2001) and addiction (Volkow et al., 2004). Our results are consistent with rodent (Laviola et al., 2003) and previous developmental imaging (Ernst et al., 2005) studies of enhanced accumbens activity during adolescence. These findings suggest that different developmental trajectories for these regions may relate to the increased impulsive and risky behaviors observed during this period of development.

Developmental changes in structure and function

Enhanced accumbens activity was paralleled by a refined pattern of activity for adolescents relative to children, but similar to adults. In contrast, adolescents showed more diffuse OFC recruitment more similar to children than adults. We interpret these data to suggest that the NAcc development may precede that of the OFC during adolescence. Protracted development of

prefrontal regions, with a transition from diffuse to focal recruitment is consistent with MRI-based neuroanatomical (Sowell et al., 1999; 2003; Gogtay et al., 2004) and fMRI studies (Casey et al., 1997, 2002; Brown et al., 2005, Durston et al., 2006) of prefrontal development (Casey et al., 2005).

Developmental changes in volume of activity within frontostriatal regions (Sowell et al., 1999) are interesting in light of known developmental processes (e.g., dendritic arborization, synaptic pruning, myelination) occurring during this period. However, neither fMRI nor MRI provide a level of analysis with which to definitively characterize the mechanism of such changes. The volume measures were used in part to constrain the interpretation of magnitude differences, but we can only speculate that our changes in volume and magnitude of activity of the NAcc and OFC reflect a fine-tuning of this circuitry with experience and development.

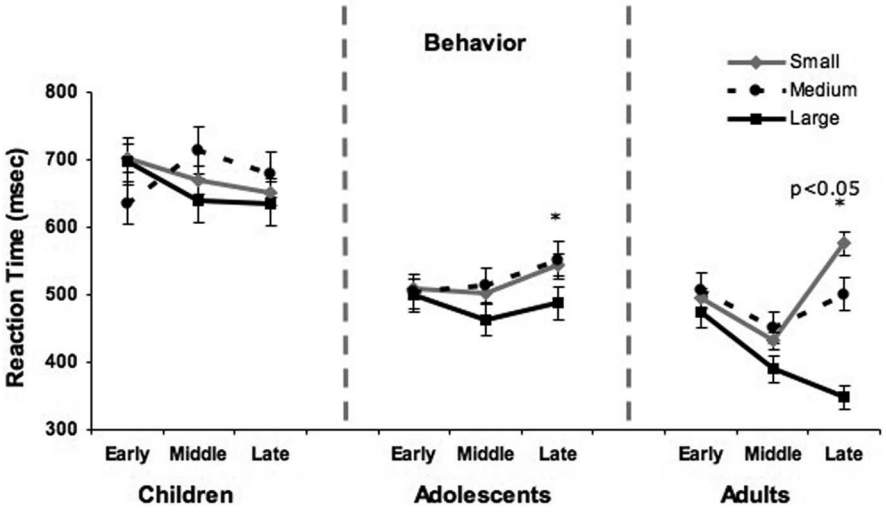


Figure 8. Behavioral results. Mean reaction time as a function of small, medium, and large reward values are shown for early, middle, and late trials of the experiment for each age group. Error bars indicate SEM. Asterisks denote slower reaction time to small and medium rewards relative to small in adolescents and slower reaction time to small relative to medium and medium relative to large in adults.

Differential recruitment of frontostriatal regions has been reported across several developmental fMRI studies (Casey et al., 2002; Monk et al., 2003; Thomas et al., 2004). Typically, these findings have been interpreted in terms of immature prefrontal regions rather than an imbalance between prefrontal and subcortical regions. Given evidence of prefrontal regions in guiding appropriate actions in different contexts (Miller and Cohen, 2001) immature prefrontal activity might hinder appropriate estimation of future outcomes and appraisal of risky choices, and might thus be less influential on reward valuation than the accumbens. This pattern is consistent with previous research showing elevated subcortical, relative to cortical activity when decisions are biased by immediate over long-term gains (McClure et al., 2004). Furthermore, accumbens activity has been shown to positively correlate with subsequent risk-taking behaviors (Kuhnen and Knutson, 2005).

Reward-related learning across development

One goal of this study was to characterize reward learning across development. Adults showed behavioral distinction to the three cues, with fastest responses to the large reward cue. Adolescents showed less discrete responses and children show little to no learning. Slower learning across development parallels the imaging results of protracted OFC development that may hinder associative learning between predictive events and reward outcomes. This interpretation is supported by animal (Hikosaka and Watanabe, 2000; Chudasama and Robbins, 2003; Cetin et al., 2004; Hosokawa et al., 2005) and human imaging (Elliott et al., 2000; O'Doherty et al., 2003; McClure et al., 2004; Cox et al., 2005; Galvan et al., 2005) studies showing the role of the OFC in learning and representing links between predictive events (stimuli and responses) and reward outcomes in optimizing choice behavior.

Few imaging studies of reward to date have been able to show differences in behavior as a function of reward outcome (Haruno et al., 2004; Delgado et al., 2005; Galvan et al., 2005). Here, our data suggest that reward-related neural responses influence behavioral output. Minimal behavioral variability might have precluded previous authors from determining whether different reward conditions bias behavioral output. One reason we were able to tease apart behavioral differences might be because our paradigm was designed to maximize behavioral responses and learning by using a continuous reinforcement schedule (Dickinson and Mackintosh, 1978; Gottlieb, 2004, 2005). Animal studies show faster learning with continuous relative to intermittent reinforcement schedules (Gottlieb, 2004) that may have explained the faster responses to large reward trials across subjects and the distinct pattern of behavior for each reward value in the adults by late trials.

Rewards are relative for different contexts and ages

Reward preference varies based on the reward context (Tversky and Kahneman, 1981; Tremblay and Schultz, 1999). Evidence from our study supports the notion that relative reward preference is exaggerated during adolescence: adolescents showed an enhanced accumbens response to the large reward and a decrease in activity to the small reward relative to other rewards and to other ages. Adolescents report greater intensity of positive feelings and more positive BOLD signal intensity than adults during a win condition (Ernst et al., 2005). The adolescents may have viewed the small reward as an omission of reward, similar to lack of an expected event at a given time, previously shown to decrease striatal activity (Davidson et al., 2004). This finding corresponded to a slowing of reaction time from early to late trials for the smaller rewards, providing additional evidence that this condition may have been perceived as more negative for adolescents. Together, these findings imply that reward perception might be influenced by changes in neural systems during adolescence (Irwin, 1993).

Developmental changes may parallel changes with learning

Recently, Pasupathy and Miller (2005) showed that, in monkeys, striatal areas detected reward contingencies first, which then seemed to bias prefrontal regions into taking action. Other work has shown that the OFC seems to be implicated in linking responses with outcomes (Elliott et al., 2000; Galvan et al., 2005). However, this effect may be dependent on maturity of prefrontal systems and reciprocal connections between frontostriatal re-

gions (Haber, 2003) that associate actions with outcome, because children and adolescents did not show learning, as indexed by mean reaction time, to the extent that adults did. It remains an open question whether the children could not learn to discriminate between the different reward values or whether they were just as happy with a small reward as a large reward.

The findings of less sensitivity in the behavioral response than in the neural response in the younger subjects may be consistent with previous learning studies showing that neural changes precede behavioral changes (Tremblay et al., 1998). The adolescents were significantly faster to larger reward trials by the end of the experiment relative to the other reward values, but the accumbens showed distinct patterns of activity to each reward value similar to adults. If this explanation were true, we might expect with additional training that the adolescents' behavioral performance would ultimately parallel the accumbens activity. Likewise, one would expect similar patterns to emerge in children, but with more extensive training.

Contrasts between current and previous findings

Although the exaggerated accumbens response in adolescents replicates those of May et al. (2004) and Ernst et al. (2005), Bjork et al. (2004) found decreased accumbens activity relative to adults during a gain versus no gain contrast. Whereas Bjork et al. (2004) reported MR signal change over the entire experiment, we examined MR changes across the entire experiment and also during early and late trials, with the later trials showing greater activation in adolescents relative to adults.

A second difference in the current study, relative to the existing literature (O'Doherty et al., 2001, Elliott et al., 2003, Galvan et al., 2005), was the lack of a main effect of reward value in the OFC across subjects. In examining this main effect, we collapsed OFC activity across age groups and across the experiment. Other reward studies of the OFC have not included developmental populations, who have diffuse and more variable patterns of activity in this region (Casey et al., 1997). Inclusion of developmental populations thus increased the variability in recruitment of this region, with less consistent patterns of OFC activity. Furthermore, our data showed that, for later trials of the experiment, OFC activity differed for larger relative to smaller rewards, but showed a less precise mapping to reward value relative to the NAcc, which showed discrete patterns of activity to each reward value across age groups, consistent with our previous work (Galvan et al., 2005) and that of others (Elliott et al., 2003).

Implications

Our results suggest that there are protracted maturational changes in top–down control systems relative to subcortical regions implicated in appetitive behaviors. These different developmental trajectories may contribute to suboptimal choices in adolescents being driven more by appetitive systems than control systems (Spear, 2000). Understanding the development of structural and functional connectivity of reward-related mesolimbic circuitry may further inform the field on the neurobiological basis of increased reward-seeking and adolescent-onset addiction.

A neural framework similar to the one we propose here has been suggested to explain addiction. Accordingly, the PFC is "hijacked" by an impulsive subcortical system, which might render it unable to appropriately modulate decisions in the context of future consequences (Bechara, 2005). Our findings are consistent with this speculation but occur during typical development. Thus, disproportionate contributions of subcortical systems relative to prefrontal regulatory systems may underlie poor decision-making that predisposes adolescents to drug use and, ultimately, addiction.

References

Bechara A (2005) Decision-making, impulse control and loss of willpower to resist drugs: a neurocognitive perspective. Nat Neurosci 8:1458–1463.

Bjork JM, Knutson B, Fong GW, Caggiano DM, Bennett SM, Hommer DW (2004) Incentive-elicited brain activation in adolescents: similarities and differences from young adults. J Neurosci 24:1793–1802.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221.

Brown TT, Lugar HM, Coalson RS, Miezin FM, Petersen SE, Schlaggar BL (2005) Developmental changes in human cerebral functional organization for word generation. Cereb Cortex 15:275–290.

Bunge S, Dudukovic NM, Thomason ME, Vaidya CJ, Gabrieli JD (2002) Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. Neuron 33:301–311.

Burgund ED, Kang HC, Kelly JE, Buckner RL, Snyder AZ, Petersen SE, Schlaggar BL (2002) The feasibility of a common stereotactic space for children and adults in fMRI studies of development. NeuroImage 17:184–200.

Casey BJ, Trainor R, Orendi J, Schubert A, Nystrom LE, Cohen JD, Noll D, Giedd J, Castellanos X, Haxby J, Forman SD, Dahl RE, Rapoport JL (1997) A pediatric functional MRI study of prefontal activation during performance of a Go-No-Go task. J Cogn Neurosci 9:835–847.

Casey BJ, Thomas KM, Davidson MC, Kunz K, Franzen P (2002) Dissociating striatal and hippocampal function developmentally with a stimulus-response compatibility task. J Neurosci 22:8647–8652.

Casey BJ, Galvan A, Hare TA (2005) Changes in cerebral functional organization during cognitive development. Curr Opin Neurobiol 15:239–244.

Cetin T, Freudenberg F, Fuchtemeier M, Koch M (2004) Dopamine in the orbitofrontal cortex regulates operant responding under a progressive ratio of reinforcement in rats. Neurosci Lett 370:114–117.

Chudasama Y, Robbins TW (2003) Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: further evidence for the functional heterogeneity of the rodent frontal cortex. J Neurosci 23:8771–8780.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Cox SM, Andrade A, Johnsrude IS (2005) Learning to like: a role for human orbitofrontal cortex in conditioned reward. J Neurosci 25:2733–2740.

Cromwell HC, Schultz W (2003) Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. J Neurophysiol 89:2823–2838.

Davidson MC, Horvitz JC, Tottenham N, Fossella JA, Watts R, Ulug A, Casey B (2004) Differential cingulate and caudate activation following unexpected nonrewarding stimuli. NeuroImage 23:1039–1045.

Delgado MR, Miller M, Inati S, Phelps EA (2005) An fMRI study of reward-related probability learning. NeuroImage 24:862–873.

Dickinson A, Mackintosh N (1978) Classical conditioning in animals. Annu Rev Psychol 29:587–612.

Durston S, Davidson MC, Tottenham N, Galvan A, Spicer J, Fossella JA, Casey B (2006) A shift from diffuse to focal cortical activity with development. Dev Sci 9:1–8.

Duvernoy HM (1999) The human brain: surface, three-dimensional sectional anatomy with MRI and blood supply, Ed 2. New York: Springer.

Elliott R, Dolan RJ, Frith CD (2000) Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies. Cereb Cortex 10:308–317.

Elliott R, Newman JL, Longe OA, Deakin JFW (2003) Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: a parametric functional magnetic resonance imaging study. J Neurosci 23:303–307.

Ernst M, Nelson EE, Jazbec SP, McClure EB, Monk CS, Leibenluft E, Blair J, Pine DS (2005) Amygdala and nucleus accumbens in responses to receipt and omission of gains in adults and adolescents. NeuroImage 25:1279–1291.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Minton MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic res-

onance imaging (fMRI): use of a cluster-size threshold. Magn Reson Med 33:636–647.
Galvan A, Hare TA, Davidson M, Spicer J, Glover G, Casey BJ (2005) The role of ventral frontostriatal circuitry in reward-based learning in humans. J Neurosci 25:8650–8656.
Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, Paus T, Evans AC, Rapoport JL (1999) Brain development during childhood and adolescence: a longitudinal MRI study. Nat Neurosci 2:861–863.
Glover GH, Thomason ME (2004) Improved combination of spiral-in/out images for BOLD fMRI. Magn Reson Med 51:863–868.
Gogtay N, Giedd JN, Lusk L, Hayashi KM, Greenstein D, Vaituzis AC, Nugent III TF, Herman DH, Clasen LS, Toga AW, Rapoport JL, Thompson PM (2004) Dynamic mapping of human cortical development during childhood through early adulthood. Proc Natl Acad Sci USA 101:8174–8179.
Gottlieb D (2004) Acquisition with partial and continuous reinforcement in pigeon autoshaping. Learn Behav 32:321–334.
Gottlieb D (2005) Acquisition with partial and continuous reinforcement in rat magazine approach. J Exp Psychol Anim Behav Process 31:319–333.
Haber SN (2003) The primate basal ganglia: parallel and integrative networks. J Chem Neuroanat 26:317–330.
Haruno M, Kuroda T, Doya K, Toyama K, Kimura M, Samejima K, Imamizu H, Kawato M (2004) A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. J Neurosci 24:1660–1665.
Hikosaka K, Watanabe M (2000) Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. Cereb Cortex 10:263–271.
Hosokawa T, Kato K, Inoue M, Mikami A (2005) Correspondence of cue activity to reward activity in the macaque orbitofrontal cortex. Neurosci Lett 389:121–182.
Hyman SE, Malenka RC (2001) Addiction and the brain: the neurobiology of compulsion and its persistence. Nat Rev Neurosci 2:695–703.
Irwin J (1993) Adolescence and risk taking: how are they related. Newbury Park, CA: Sage.
Kang HC, Burgund ED, Lugar HM, Petersen SE, Schlaggar BL (2003) Comparison of functional activation foci in children and adults using a common stereotactic space. NeuroImage 19:16–28.
Knutson B, Adams CM, Fong GW, Hommer D (2001) Anticipation of increasing monetary reward selectively recruits nucleus accumbens. J Neurosci 21:1–5.
Kuhnen CM, Knutson B (2005) The neural basis of financial risk taking. Neuron 47:763–770.
Laviola G, Adriani W, Terranova ML, Gerra G (1999) Psychobiological risk factors for vulnerability to psychostimulants in human adolescents and animal models. Neurosci Biobehav Rev 23:993–1010.
Laviola G, Macri S, Morley-Fletcher S, Adriani W (2003) Risk-taking behavior in adolescent mice: psychobiological determinants and early epigenetic influence. Neurosci Biobehav Rev 27:19–31.
Luna B, Sweeney JA (2004) The emergence of collaborative brain function: fMRI studies of the development of response inhibition. Ann NY Acad Sci 1021:296–309.
Luna B, Thulborn KR, Munoz DP, Merriam EP, Garver KE, Minshew NJ, Keshavan MS, Genovese CR, Eddy WF, Sweeney JA (2001) Maturation of widely distributed brain function subserves cognitive development. NeuroImage 13:786–793.
May JC, Delgado MR, Dahl RE, Stenger A, Ryan ND, Fiez JA, Carter CS (2004) Event-related functional magnetic resonance imaging of reward-related brain circuitry in children and adolescents. Biol Psychiatry 55:359–366.
McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate neural systems value immediate and delayed monetary rewards. Science 306:503–507.
Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202.
Monk CS, McClure EB, Nelson EE, Zarahn E, Bilder RM, Leibenluft E, Charney DS, Ernst M, Pine DS (2003) Adolescent immaturity in attention-related brain engagement to emotional facial expressions. NeuroImage 20:420–428.
Moses P, Roe K, Buxton RB, Wong EC, Frank LR, Stiles J (2002) Functional MRI of global and local processing in children. NeuroImage 16:415–424.
O'Doherty J, Kringelbach M, Rolls ET, Hornak J, Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci 4:95–102.
O'Doherty J, Critchley H, Deichmann R, Dolan RJ (2003) Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. J Neurosci 23:7931–7939.
Pasupathy A, Miller EK (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature 433:873–876.
Pecina S, Cagniard B, Berridge KC, Aldridge JW, Zhuang X (2003) Hyperdopaminergic mutant mice have higher "wanting" but not "liking" for sweet rewards. J Neurosci 23:9395–9402.
Rosenberg D, Lewis DA (1994) Changes in the dopaminergic innervation of monkey prefrontal cortex during late postnatal development: a tyrosine hydroxylase immunohistochemical study. Biol Psychiatry 36:272–277.
Rosenberg D, Lewis DA (1995) Postnatal maturation of the dopaminergic innervation of monkey prefrontal and motor cortices: a tyrosine hydroxylase immunohistochemical analysis. J Comp Neurol 358:383–400.
Rubia K, Overmeyer S, Taylor E, Brammer M, Williams S, Simmons A, Andrew C, Bullmore E (2000) Functional frontalisation with age: mapping neurodevelopmental trajectories with fMRI. Neursci Biobehav Rev 24:13–19.
Silveri MM, Tzilos GK, Pimentel PJ, Yurgelun-Todd DA (2004) Trajectories of adolescents emotional and cognitive development: effects of sex and risk for drug use. Ann NY Acad Sci 1021:363–370.
Sowell ER, Thompson PM, Holmes CJ, Jernigan TL, Toga AW (1999) In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. Nat Neurosci 2:859–861.
Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW (2003) Mapping cortical change across the human life span. Nat Neurosci 6:309–315.
Spear LP (2000) The adolescent brain and age-related behavioral manifestations. Neurosci Biobehav Rev 24:417–463.
Steinberg L (2004) Risk taking in adolescence: what changes, and why? Ann NY Acad Sci 1021:51–58.
Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme.
Thomas KM, Hunt R, Vizueta N, Sommer T, Durston S, Yang Y, Worden MS (2004) Evidence of developmental differences in implicit sequence learning: an fMRI study of children and adults. J Cogn Neurosci 16:1339–1351.
Tremblay K, Kraus N, McGee T (1998) The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. NeuroReport 9:3557–3560.
Tremblay L, Schultz W (1999) Relative reward preference in primate orbitofrontal cortex. Nature 398:704–708.
Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. Science 211:453–458.
Volkow ND, Fowler J, Wang G, Swanson JM (2004) Dopamine in drug abuse and addiction: results from imaging studies and treatment implications. Mol Psychiatry 9:557–569.
Zald DH, Boileau I, El-Dearedy W, Gunn R, McGlone F, Dichter GS, Dagher A (2004) Dopamine transmission in the human striatum during monetary reward tasks. J Neurosci 24:4105–4112.

11. Y. Montet *et al*., *Nature Struct. Biol.* **4**, 523 (1997).
12. L. Holm, C. Sander, *Proteins* **19**, 165 (1994).
13. T. Shanmugasundaram, G. K. Kumar, H. G. Wood, *Biochemistry* **27**, 6499 (1988).
14. S. Menon, S. W. Ragsdale, *J. Biol. Chem.* **274**, 11513 (1999).
15. S. W. Ragsdale, J. E. Clark, L. G. Ljungdahl, L. L. Lundie, H. L. Drake, *J. Biol. Chem.* **258**, 2364 (1983).
16. S. W. Ragsdale, in *Enzyme-Catalyzed Electron and Radical Transfer*, A. Holzenburg, N. Scrutton, Eds. (Plenum, New York, 2000), vol. 35, pp. 487–518.
17. J. W. Peters, W. N. Lanzilotta, B. J. Lemon, L. C. Seefeldt, *Science* **282**, 1853 (1998).
18. Y. Nicolet, C. Piras, P. Legrand, C. E. Hatchikian, J. C. Fontecilla-Camps, *Structure* **7**, 13 (1999).
19. W. K. Russell, C. M. V. Stalhandske, J. Q. Xia, R. A. Scott, P. A. Lindahl, *J. Am. Chem. Soc.* **120**, 7502 (1998).
20. S.-I. Hu, H. L. Drake, H. G. Wood, *J. Bacteriol.* **149**, 440 (1982).
21. M. Kumar, S. W. Ragsdale, *J. Am. Chem. Soc.* **114**, 8713 (1992).
22. C. Harford, B. Sarkar, *Acc. Chem. Res.* **30**, 123 (1997).
23. W. Huang *et al*., *Structure* **5**, 691 (1997).
24. S. Nagashima *et al*., *Nature Struct. Biol.* **5**, 347 (1998).
25. U. Ermler, W. Grabarse, S. Shima, M. Goubeaud, R. K. Thauer, *Science* **278**, 1457 (1997).
26. R. K. Thauer, *Microbiology UK* **144**, 2377 (1998).
27. D. P. Barondeau, P. A. Lindahl, *J. Am. Chem. Soc.* **119**, 3959 (1997).
28. S. Jaron, N. J. Blackburn, *Biochemistry* **38**, 15086 (1999).
29. A. Szulc, D. Meyerstein, H. Cohen, *Inorg. Chim. Acta* **270**, 440 (1998).
30. E. L. Maynard, C. Sewell, P. A. Lindahl, *J. Am. Chem. Soc.* **123**, 4697 (2001).
31. J. Seravalli, M. Kumar, S. W. Ragsdale, *Biochemistry* **41**, 1807 (2002).
32. J. R. Andreese, A. Schaupp, C. Neuraute, A. Brown, L. G. Ljungdahl, *J. Bacteriol.* **114**, 743 (1973).
33. Z. Otwinowski, W. Minor, *Methods Enzymol.* **276**, 307 (1997).
34. I. Uson, G. M. Sheldrick, *Curr. Opin. Struct. Biol.* **9**, 643 (1999).
35. A. T. Brünger *et al*., *Acta Crystallogr.* **D54**, 905 (1998).
36. C. R. Kissinger, D. K. Gehlhaar, D. B. Fogel, *Acta Crystallogr.* **D55**, 484 (1999).
37. K. Cowtan, *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34 (1994).
38. K. D. Cowtan, K. Y. Zhang, *Prog. Biophys. Mol. Biol.* **72**, 245 (1999).
39. D. E. McRee, *J. Struct. Biol.* **125**, 156 (1999).
40. T. A. Jones, J.-Y. Zou, S. W. Cowen, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991).
41. R. A. Laskowski, M. W. McArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).
42. M. Carson, *Methods Enzymol.* **277**, 493 (1997).
43. J. A. Christopher, thesis, Texas A&M University (1998).
44. M. L. Youngblood, D. W. Margerum, *Inorg. Chem.* **19**, 3068 (1980).
45. P. Strop *et al*., *Biochemistry* **40**, 651 (2001).

46. We thank T. Earnest and G. McDermott at Advanced Light Source (ALS) for help with data collection. Support has been provided by the Templeton Foundation (C.L.D.) and National Institutes of Health grant R01-GM39451 (S.W.R.). The data collection facilities at ALS are funded by the U.S. Department of Energy, Office of Basic Energy Sciences.

8 July 2002; accepted 3 September 2002

Shadows Cast by Retinal Blood Vessels Mapped in Primary Visual Cortex

Daniel L. Adams* and Jonathan C. Horton

The mammalian eye is a remarkable optical device, but its design is not perfect. The blood vessels that supply the inner retina are located in front of the photoreceptor layer, blocking access to light. Their shadows create a pattern of blindness in the field of vision that corresponds precisely to the location of the largest vessels in the eye. We show here that in squirrel monkeys, focal deprivation by blood vessels leads to rewiring of the eye's geniculocortical projections, imprinting an image of the retinal vascular tree onto the primary visual cortex. This process illustrates vividly that local imbalances in neuronal activity can influence column formation during normal development.

In many higher species, the geniculate afferents serving right and left eyes are segregated in the brain in layer 4 of primary (striate, V1) visual cortex into a mosaic of interdigitated inputs called ocular dominance columns (*1*). The formation of these columns has been scrutinized in the hope of elucidating the processes that guide the development of the cerebral cortex (*2*). Some experiments show an essential role for neuronal activity, because infusion of tetrodotoxin into both eyes seems to prevent the emergence of ocular dominance columns (*3*). According to this idea, the correlated discharge of neighboring ganglion cells, perhaps generated by spontaneous waves of retinal activity (*4*, *5*), induces local cortical imbalances in the density of geniculate afferents driven by each eye. These small imbalances may be amplified by competitive synaptic interactions between geniculocortical afferents to generate a mature system of ocular dominance columns (*6*–*9*).

Doubt has been cast on this explanation for column development by recent experiments showing that ocular dominance columns appear in ferrets despite early bilateral enucleation (*10*). This finding suggests that column formation is ordained by intrinsic molecular cues, and not by patterns of neuronal activity. Although neuronal activity has a disputed role in column induction, it certainly can influence column morphology at a later stage in development. Raising a newborn animal with monocular eyelid suture, to mimic congenital cataract, causes shrinkage of the deprived eye's columns and expansion of the normal eye's columns (*11*–*14*). Here, an imbalance in neuronal activity has a profound influence on the shape and size of the ocular dominance columns. The relevance of these experiments, however, can be questioned on the grounds that suture of the eyelids is an artificial manipulation, creating an extreme imbalance in the stimulation of each retina. It has not been shown that under natural circumstances, competition between the two eyes has an important influence on the appearance of the ocular dominance columns.

Cortical representation of retinal vessels. We now show that visual deprivation occurs routinely in normal primates, affecting ocular dominance patterns. Helmholtz discovered while plotting his blind spot that he could discern the proximal stumps of large blood vessels emerging from the optic disc (*15*). With refinement of perimetric techniques, subsequent investigators have plotted these so-called angioscotomas over wide portions of the visual field (*16*, *17*). The hemoglobin in red blood cells absorbs light, casting a shadow from blood vessels in the inner retina onto photoreceptors located underneath. We have found a representation of angioscotomas in striate cortex of the squirrel monkey (*Saimiri sciureus*).

The squirrel monkey was reported originally to lack ocular dominance columns (*18*, *19*). They are present, in fact, but often appear poorly developed and irregular (*20*). To examine their pattern more closely, we enucleated one eye and stained striate cortex for the mitochondrial enzyme cytochrome oxidase (CO) (*21*–*23*). The ocular dominance columns are revealed by this approach because metabolic activity is reduced in the missing eye's columns. After perfusion, striate cortex was stripped from the white matter, flattened in a single piece, and sectioned tangentially to the surface. Flatmount preparations enhance the likelihood of detecting subtle metabolic patterns, such as the optic disc representation, which lies buried in the calcarine sulcus (*24*).

Beckman Vision Center, University of California, San Francisco, San Francisco, CA 94143–0730, USA.

*To whom correspondence should be addressed at Beckman Vision Center, University of California, San Francisco, 10 Kirkham Street, San Francisco, CA 94143–0730, USA. E-mail: dadams@itsa.ucsf.edu

Downloaded from www.sciencemag.org on November 14, 2015

Ocular dominance is expressed most strongly in layer 4C of striate cortex. CO montage of this layer revealed a fuzzy pattern of fine ocular dominance columns (Fig. 1). The left eye's optic disc representation appeared as a large, homogeneous zone of CO activity in the right cortex. It stained darkly, because the left eye was enucleated in this animal. In the periphery, a large pale zone corresponded to the representation of the far nasal retina of the left eye (monocular crescent).

Radiating from the optic disc representation were more than a dozen, thin, dark profiles. They represented the blood vessels emanating from the optic disc of the left eye. This was confirmed by matching each element in the cortical pattern to its corresponding retinal blood vessel (Fig. 1, C and D).

How do angioscotoma representations form? Angioscotoma representations cannot form before birth, because the uterine environment is dark. We propose that after birth (Fig. 2A), the right cortex, for instance, receives binocular input organized into nascent columns, except at the representation of the left optic disc, where input comes exclusively from the right eye. Once the baby monkey opens its eyes (Fig. 2B), light enters through the pupil and shadows are cast by retinal blood vessels. Ganglion cells in the left eye that receive input from photoreceptors beneath blood vessels are not driven effectively by visual stimulation. In turn, these ganglion cells fail to activate lateral geniculate neurons. Consequently, the axon terminals of these geniculate fibers retract in striate cortex, and the territory they relinquish is taken over by afferents serving the right eye. After the critical period, geniculocortical afferents become frozen in place. When one eye is enucleated (Fig. 2C), CO reveals the ocular dominance columns, which have been rearranged to correspond to the cortical representation of retinal blood vessels.

To confirm this hypothesis, we double-labeled the ocular dominance columns in this animal (and two others) by injecting an anatomical tracer, [^{3}H]proline, into the remaining right eye (*25*, *26*). The proline and CO patterns were identical (compare Figs. 1A and 3). As expected, the angioscotoma representations in striate cortex were solidly labeled, as intensely as the blind spot representation (Fig. 2C). This experiment proved that cortical angioscotoma representations are generated by anatomical rearrangement of geniculocortical inputs. It should be emphasized that in three control animals, with intact vision in both eyes, no CO pattern was visible in layer 4C. Removal of one eye is required to reveal the cortical angioscotoma representations with CO.

Angioscotomas were represented in striate cortex of 9 of 12 squirrel monkeys. Shadows generated in the contralateral eye

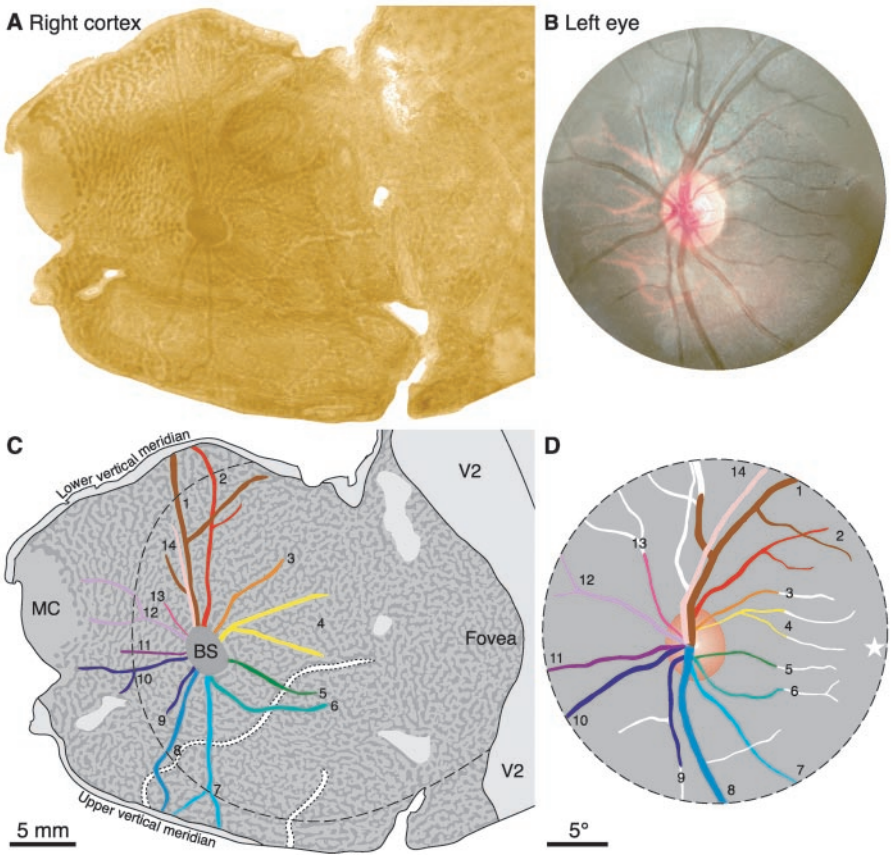


Fig. 1. Representation of angioscotomas in striate cortex. (**A**) Flatmount montage stained for CO after removal of the left eye to reveal the ocular dominance columns. The dark lines emerging from the central dark oval correspond to the blood vessels of the left eye. (**B**) Fundus photograph showing optic disc and retinal blood vessels. (**C**) Drawing of vessel representations, color-coded to match the retinal drawing. Dashed line denotes circumference of retinal photograph in (B), projected onto the cortex. MC, monocular crescent; BS, blind spot. (**D**) Drawing of retina, with represented portions of vessels coded by color. Star denotes the fovea.

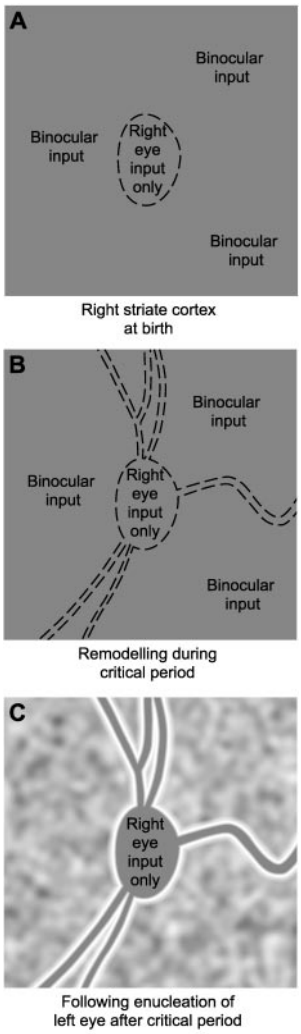


Fig. 2. CO staining of angioscotoma representations. (**A**) At birth, CO staining is uniform. Angioscotomas are absent, but the right eye already monopolizes the left eye's blind spot representation. (**B**) After light exposure, vessels begin to cast shadows, generating angioscotoma representations. Their pattern becomes immutable at the end of the critical period. (**C**) Removal of one eye induces an undulating pattern of CO activity, corresponding to the ocular dominance columns. The angioscotomas appear silhouetted against this textured background, because they stain solidly. Note the tendency for the other eye to "frost" the vessel and blind spot representations.

predominated, because blood vessels are generally larger in nasal retina, especially near the optic disc. A few vessel representations, however, could be traced across the vertical meridian (V1/V2 border) and into the opposite cortex (Figs. 3 and 4). Their contrast was reversed, compared with other angioscotomas in the same cortex, because they originated from the temporal retina of the ipsilateral eye.

Each angioscotoma representation was fringed by cortex that exhibited opposite CO contrast (Fig. 5). This "frosting" enhanced the visibility of the angioscotoma pattern by highlighting each vessel representation. Apparently, when retraction of geniculate afferents occurs during the critical period, the displaced afferents form a solid phalanx along the edges of the representation of each angioscotoma. The displaced afferents dominate in these flanking zones, expelling those from the other eye. This is noteworthy, because neither eye enjoys any obvious competitive advantage in visual stimulation right next to an angioscotoma.

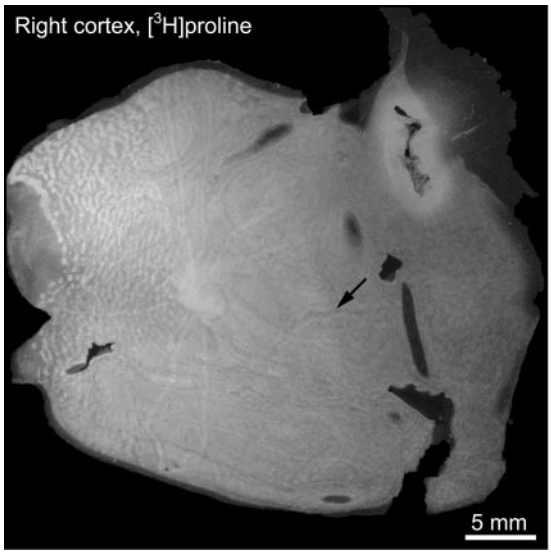


Fig. 3. Autoradiograph prepared after tracer injection into the remaining right eye, showing label (bright regions) distributed in a pattern identical to that in Fig. 1A. The congruence of the patterns confirms that CO staining after monocular enucleation precisely reflects the distribution of geniculocortical inputs. The angioscotomas and the blind spot are solidly labeled, as predicted in Fig. 2C. Arrow denotes an unlabeled angioscotoma, arising from a temporal vessel in the right eye.

Precision of the visual field representation. The biggest vessels represented in the cortex were retinal veins near the optic disc. They attained a diameter of nearly 100 μm (0.62°). The smallest were vessels approaching the fovea, less than 30 μm wide (0.17°). To calculate shadow widths, we measured exit pupil diameter, distance from exit pupil to retinal vessel, retinal vessel diameter, and distance from retinal vessel to photoreceptors (*27*, *28*). Many smaller vessels failed to eclipse the pupil completely and therefore produced no umbra. For example, a 28-μm-diameter vessel 7.5° from the fovea cast only penumbra, measuring 83 μm in width (Fig. 6). From an expression for cortical magnification factor along an isoeccentricity ring

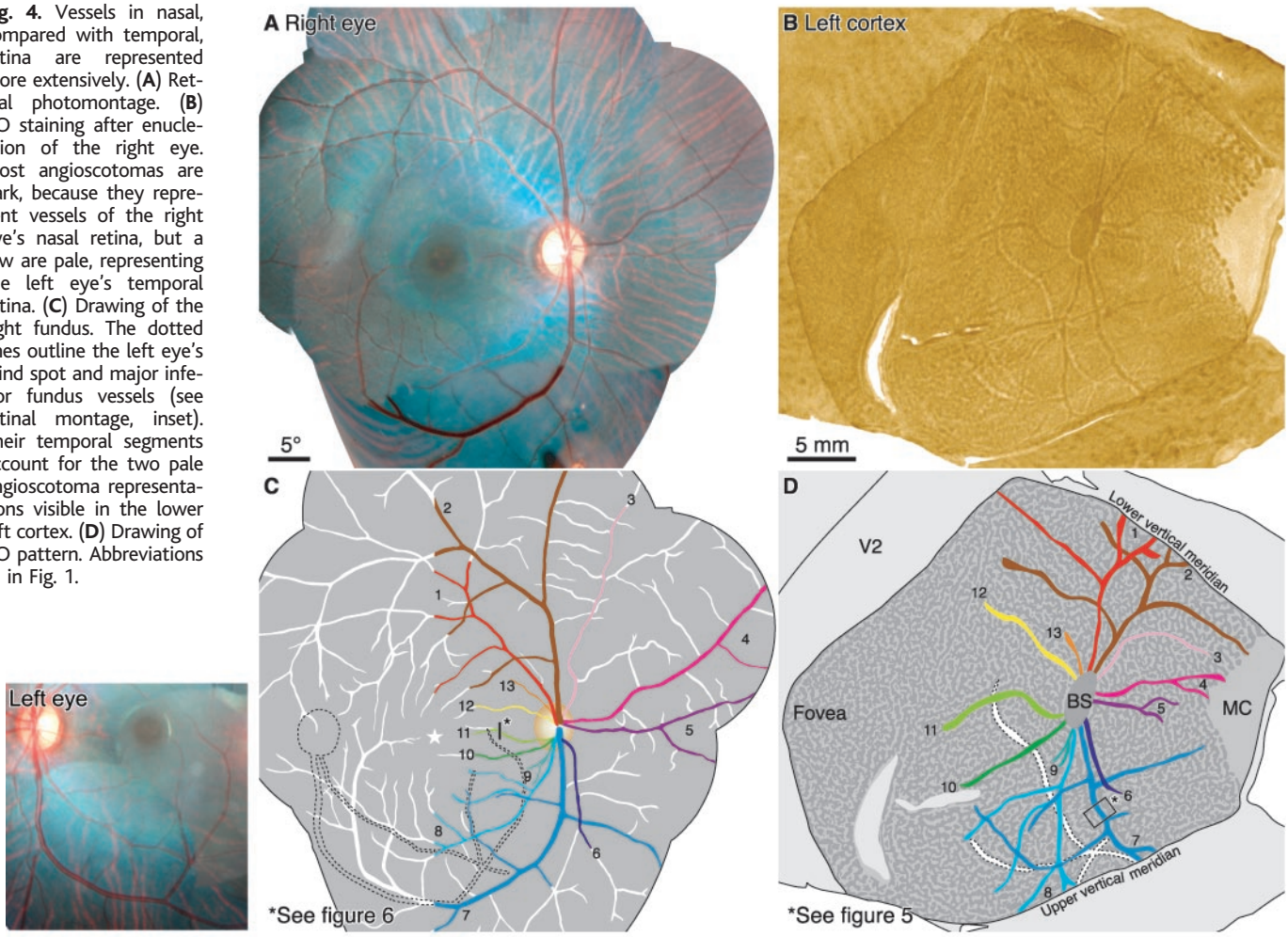


Fig. 4. Vessels in nasal, compared with temporal, retina are represented more extensively. (**A**) Retinal photomontage. (**B**) CO staining after enucleation of the right eye. Most angioscotomas are dark, because they represent vessels of the right eye's nasal retina, but a few are pale, representing the left eye's temporal retina. (**C**) Drawing of the right fundus. The dotted lines outline the left eye's blind spot and major inferior fundus vessels (see retinal montage, inset). Their temporal segments account for the two pale angioscotoma representations visible in the lower left cortex. (**D**) Drawing of CO pattern. Abbreviations as in Fig. 1.

(derived below), we calculated that an 83-μm shadow should produce a cortical representation 737 μm wide. In fact, the vessel's cortical profile measured only 350 μm (see vessel 11, Fig. 4D), indicating that just the central ~50% of the shadow was sufficiently dense to induce remodeling of geniculocortical afferents. This portion of the shadow was only three to four cones wide.

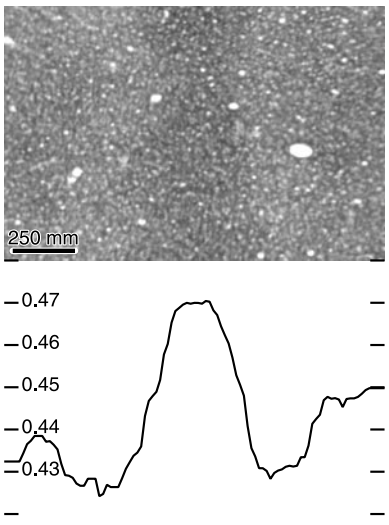


Fig. 5. Vessel representations are flanked by CO activity of opposite contrast. A vertically oriented dark angioscotoma representation (Fig. 4D) is surrounded by pale staining. Mean optical density of each column of pixels is plotted below.

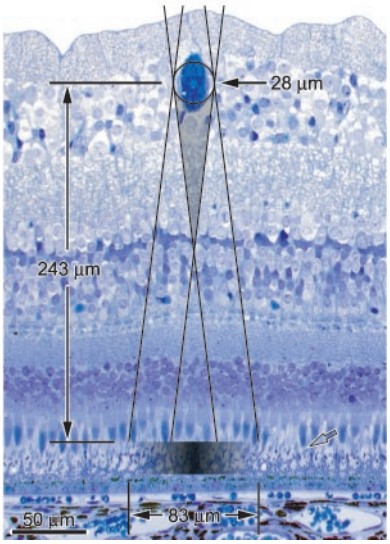


Fig. 6. Optics of shadow formation. Toluidine blue–stained retinal cross section (Fig. 4C), showing the shadow cast by a blood vessel in the ganglion cell layer with a 28-μm lumen diameter, when fully inflated. Pairs of lines, projected from opposite margins of a 1.9-mm pupil located 8.4 mm in front of the retina, show the boundaries of the shadow obscuring photoreceptors. The umbra (gray triangle) does not reach the photoreceptor layer in this example, because the vessel is too small. Gradient indicates penumbra, which reached a maximum optical density of 0.45. Open arrow shows the light-recipient boundary between cone inner and outer segments.

Generation of a retinotopic map. The angioscotoma representations provided a simple means for construction of a visual field map in striate cortex (Fig. 7). First, dozens of prominent landmarks, such as vessel crossings or bifurcations, were identified in the retina and projected onto a tangent screen to determine their eccentricity. These data were used to calibrate a system of rings and rays, which was superimposed on an image of the nasal retina. This image was transformed by computer to match retinal features precisely to their cortical representation (*29*), thereby generating a retinotopic map. This map is truly retinotopic, as opposed to visuotopic, and therefore immune to errors introduced by eye movements or inaccurate plotting of receptive fields (*30*).

Areal cortical magnification factor, M_a, was derived by measuring the area (mm^2) of each compartment in the map and dividing by the area (degrees squared) of its corresponding visual field zone. The data were plotted as a function of eccentricity (Fig. 7D) for four hemispheres from two monkeys. Cortical magnification factor in the squirrel monkey was similar to the value reported in the macaque (*31*), showing that these evolutionarily divergent primate species emphasize macular vision to the same extent.

Discussion. Retinal blood vessels give rise to angioscotomas in the visual field, just as the optic disc creates a blind spot. However, the mechanism is quite different. The

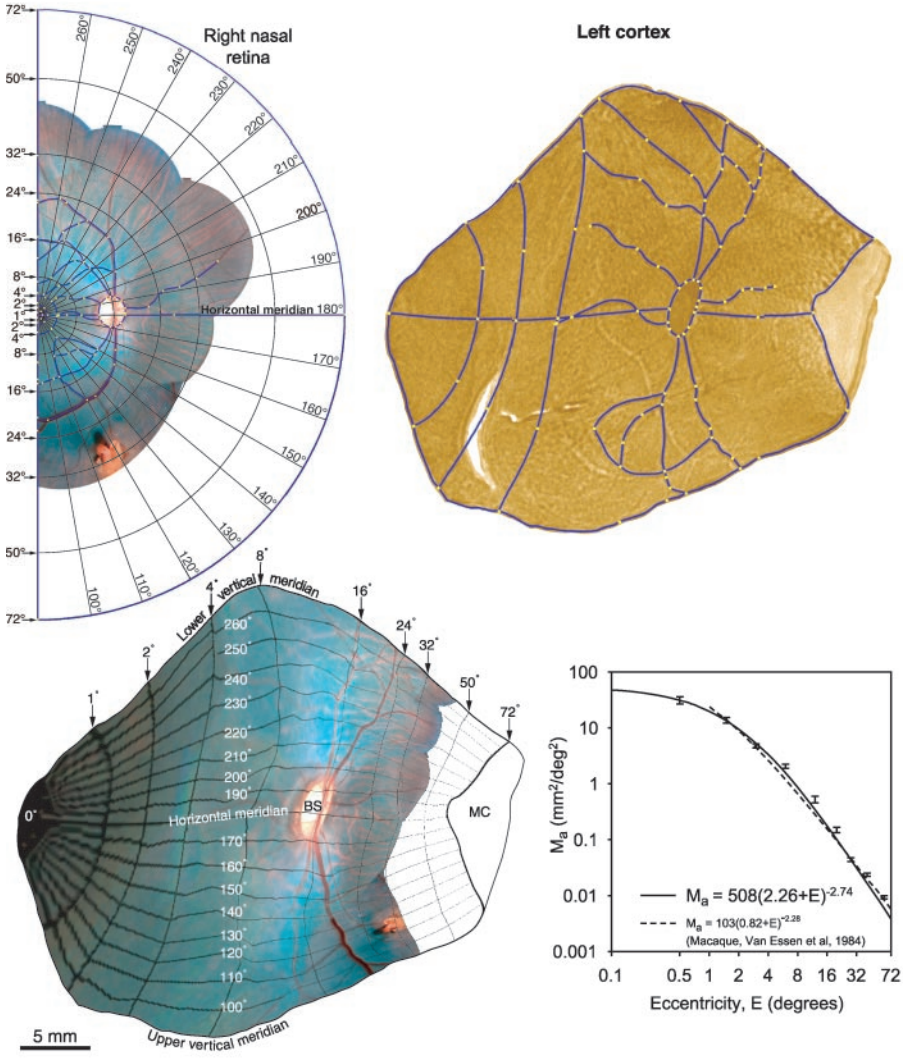


Fig. 7. Generation of a retinotopic map. Ring and ray pattern is superimposed on the right retina, using landmarks plotted in vivo. Correspondence points (yellow squares) are placed on matching features of the retina and cortex. Blue lines are placed on segments between correspondence points. They include blood vessels, vertical meridian (V1/V2) border, horizontal meridian, and the boundary between binocular and monocular cortex (measured in vivo at 72°). Warping of the retina onto the cortex produces a retinotopic map. Areal cortical magnification factor (mean ± SEM) for four cortices is graphed versus eccentricity.

optic disc produces a hole, devoid of ganglion cells, in the fabric of the retina. The other eye fills in the cortical gap, visible as a solid oval amid a sea of ocular dominance columns. This occurs in utero, before visual experience has any influence (*32*).

Angioscotomas, on the other hand, arise because a few photoreceptors are condemned to a life of idleness owing to their location beneath blood vessels. The retina is otherwise normal. Deprivation by vascular shadows induces local remodeling of geniculate afferents, in a cortical distribution that corresponds to the trajectory of blood vessels within the retina. The process is analogous to amblyopia induced by a congenital cataract. A lens opacity, however, provokes global shrinkage of geniculate afferents in the cortex (*33*). Our results show that amblyopia can be exquisitely focal within an eye. It probably occurs underneath retinal blood vessels in primates other than squirrel monkeys, including humans.

Angioscotoma representations in the cortex provide a natural example of how visual activity can influence the pattern formed by ocular dominance columns. They are essentially nothing more than long, threadlike ocular dominance columns, generated because deprivation in one eye tips the local competition for cortical territory in favor of the other eye. The deprived eye is compensated by preferential assignment of the closest adjacent territory—as reflected by the "frosting" lining the angioscotoma representations (Fig. 5). This phenomenon presumably reflects an innate tendency to allot cortical territory in a reciprocal manner (*6*–*8*, *34*, *35*). The global periodicity of the columns resulting from this process is probably under genetic control (*36*). However, our data show that local imbalances in neuronal activity can determine the precise layout of the columns in any given region of cortex.

We have searched in vain for angioscotoma representations in striate cortex of the macaque and human. In general, these species have larger ocular dominance columns than squirrel monkeys. In the macaque, we have projected the retinal vasculature onto striate cortex, to compare the predicted size of individual vessel representations with the width of ocular dominance columns. The ocular dominance columns are much larger. We speculate that because the columns are larger, their pattern is resistant to the influence of angioscotomas during development. In the macaque, the amount of cortex deprived by angioscotomas is simply too small to compel the rearrangement of ocular dominance columns. Nonetheless, shouldn't an angioscotoma carve a narrow swath through any column, no matter how big? Apparently, this is prevented by the higher degree of segregation of geniculocortical afferents in macaques, compared with squirrel monkeys. In squirrel monkeys, the fine grain and weak segregation of the ocular dominance columns create perfect conditions for the formation of angioscotoma representations.

Amblyogenic shadows just a few cones wide were represented in the cortex, reflecting the remarkable precision of the visual field map in layer 4C (*37*). In other layers, angioscotomas were not visible, because of the coarser resolution of the visual field map and the intermingling of ocular channels. The representation of angioscotomas was most extensive at eccentricities from 4° to 30°. Within this range, an ideal balance exists between cortical magnification factor and blood vessel size. Beyond 30°, vessel size and magnification factor both decrease, so most angioscotoma representations disappear. Toward the fovea, blood vessels rapidly taper. Their reduction is offset by a steep rise in cortical magnification factor, but nevertheless, angioscotoma representations abruptly peter out at 3° to 4°. This occurs because vessels near the fovea contain too little blood to cast a sufficiently dense shadow.

Angioscotomas are rarely noticed in our vision because they are filled in perceptually, just like the blind spot. Remarkably, angioscotomas amount collectively to a greater portion of the visual field than the blind spot. Nature has struck a compromise in the design of the eye, by allowing blood vessels to obscure some photoreceptors. At least the largest vessels in the retina are prohibited from approaching the fovea, so that we are scarcely aware of their presence.

References and Notes

1. D. H. Hubel, T. N. Wiesel, *Nature* **221**, 747 (1969).
2. L. C. Katz, C. J. Shatz, *Science* **274**, 1133 (1996).
3. M. P. Stryker, W. A. Harris, *J. Neurosci.* **6**, 2117 (1986).
4. M. Meister, R. O. L. Wong, D. A. Baylor, C. J. Shatz, *Science* **252**, 939 (1991).
5. R. Mooney, A. A. Penn, R. Gallego, C. J. Shatz, *Neuron* **17**, 863 (1996).
6. C. von der Malsburg, D. J. Willshaw, *Exp. Brain Res. Suppl.* **1**, 463 (1976).
7. N. V. Swindale, *Proc. R. Soc. London Ser. B* **208**, 243 (1980).
8. K. D. Miller, J. B. Keller, M. P. Stryker, *Science* **245**, 605 (1989).
9. G. J. Goodhill, *Network* **9**, 419 (1998).
10. J. C. Crowley, L. C. Katz, *Nature Neurosci.* **2**, 1125 (1999).
11. D. H. Hubel, T. N. Wiesel, S. LeVay, *Philos. Trans. R. Soc. London Ser. B* **278**, 377 (1977).
12. C. J. Shatz, M. P. Stryker, *J. Physiol. (London)* **281**, 267 (1978).
13. N. V. Swindale, F. Vital-Durand, C. Blakemore, *Proc. R. Soc. London Ser. B* **213**, 435 (1981).
14. M. C. Crair, E. S. Ruthazer, D. C. Gillespie, M. P. Stryker, *Neuron* **19**, 307 (1997).
15. H. von Helmholtz, in *Handbook of Physiological Optics* (Optical Society of America, Menasha, WI, 1924), vol. 2, p. 28.
16. J. N. Evans, *Am. J. Ophthalmol.* **9**, 489 (1926).
17. U. Schiefer *et al.*, *Vision Res.* **39**, 1897 (1999).
18. M. S. Livingstone, *J. Neurosci.* **16**, 2086 (1996).
19. A. E. Hendrickson, J. R. Wilson, M. P. Ogren, *J. Comp. Neurol.* **182**, 123 (1978).
20. J. C. Horton, D. R. Hocking, *J. Neurosci.* **16**, 5510 (1996).
21. Experiments were performed on 12 adult squirrel monkeys from the California Regional Primate Center, Davis, CA. All procedures were approved by the University of California, San Francisco, Committee on Animal Research. Each animal was anesthetized with 2% isoflurane and paralyzed with succinylcholine (45 mg/kg, intravenously). The animal was placed in a stereotaxic frame mounted on a tripod to orient the visual axis perpendicular to a large tangent screen. Photographs were taken of both retinae and landmarks were projected onto a tangent screen with a fundus camera to pinpoint their eccentricity. Retinal magnification factor was 161 μm per degree in the central 24°. One eye was then enucleated under sterile conditions. Buprenorphine (0.03 mg/kg) was administered for 48 hours after surgery for analgesia. After a survival time of at least 10 days, the animal was killed with sodium pentobarbital (150 mg/kg) and perfused with 1% paraformaldehyde. Flatmounts were cut at 30 to 40 μm, air-dried on slides, reacted for CO, and montaged with Photoshop 6.0.
22. M. T. T. Wong-Riley, *Brain Res.* **171**, 11 (1979).
23. J. C. Horton, *Philos. Trans. R. Soc. London Ser. B* **304**, 199 (1984).
24. J. C. Horton, D. R. Hocking, *J. Neurosci.* **16**, 7228 (1996).
25. Two millicuries of L-[2,3,4,5-^{3}H]proline were reconstituted in 20 μl of saline and injected into the vitreous of one eye. Alternate cortical sections were processed for CO or autoradiography.
26. T. N. Wiesel, D. H. Hubel, D. M. K. Lam, *Brain Res.* **79**, 273 (1974).
27. R. A. Applegate, A. Bradley, W. A. van Heuven, *Invest. Ophthalmol. Vis. Sci.* **31**, 2088 (1990).
28. Entrance pupil size was measured by photography with a 100-mm macro-lens in five squirrel monkeys aged 1 week; it was 1.9 ± 0.03 mm (SEM), with a range of 1.7 to 2.0 mm. This value was used for exit pupil size, because for small pupils, the difference between entrance and exit pupil size is negligible. Distance from pupil to retina was measured in three newborn squirrel monkeys with A-mode ultrasound; it was 8.4 ± 0.12 mm (SEM). Adult retinae were sectioned at 1 μm in Epon-Araldite and the lumen diameter of identified blood vessels was measured. Distance from retinal vessel to photoreceptors was also measured in these sections. Correction was made for retinal growth and tissue shrinkage.
29. About 50 to 100 correspondence points were used to warp the nasal retina onto the cortex, using Elastic Reality 3.0 (www.avid.com). No angioscotoma representations were present centrally, so points were obtained from Cowey's microelectrode mapping of the central 4° (*30*). Linear cortical magnification factor along polar rays (M_p) and along eccentricity rings (M_e) were calculated independently: $M_p = 9.5\,(0.95 + E)^{-1.03}$ and $M_e = 188\,(6.45 + E)^{-1.85}$, where E is eccentricity.
30. A. Cowey, *J. Neurophysiol.* **27**, 366 (1964).
31. D. C. Van Essen, W. T. Newsome, J. H. Maunsell, *Vision Res.* **24**, 429 (1984).
32. J. C. Horton, D. R. Hocking, *J. Neurosci.* **16**, 1791 (1996).
33. _____, *J. Neurosci.* **17**, 3684 (1997).
34. D. G. Jones, R. C. Van Sluyters, K. M. Murphy, *J. Neurosci.* **11**, 3794 (1991).
35. G. J. Goodhill, *Biol. Cybern.* **69**, 109 (1993).
36. M. Kaschube, F. Wolf, T. Geisel, S. Löwel, *J. Neurosci.* **22**, 7206 (2002).
37. G. G. Blasdel, D. Fitzpatrick, *J. Neurosci.* **4**, 880 (1984).

38. This work was supported by the National Eye Institute (grant R01 EY10217 and core grant EY02162), That Man May See, The Bunter Fund, and a Lew W. Wasserman Merit Award from Research to Prevent Blindness. We thank D. Hocking and I. Wood for help with tissue processing and R. Troyer for assistance with animal care and surgery. We thank the California Regional Primate Research Center, which is supported by NIH Base Grant RR00169. L. Sincich, N. Swindale, and G. Westheimer provided valuable comments on the manuscript.

10 June 2002; accepted 29 July 2002

Science
AAAS

Shadows Cast by Retinal Blood Vessels Mapped in Primary Visual Cortex

Daniel L. Adams and Jonathan C. Horton

Science **298**, 572 (2002);
DOI: 10.1126/science.1074887

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of November 14, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/298/5593/572.full.html

A list of selected additional articles on the Science Web sites **related to this article** can be found at:
http://www.sciencemag.org/content/298/5593/572.full.html#related

This article **cites 32 articles**, 18 of which can be accessed free:
http://www.sciencemag.org/content/298/5593/572.full.html#ref-list-1

This article has been **cited by** 23 article(s) on the ISI Web of Science

This article has been **cited by** 12 articles hosted by HighWire Press; see:
http://www.sciencemag.org/content/298/5593/572.full.html#related-urls

This article appears in the following **subject collections:**
Development
http://www.sciencemag.org/cgi/collection/development

Downloaded from www.sciencemag.org on November 14, 2015

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2002 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

Clinical and Translational Report

Cell Metabolism

A Smartphone App Reveals Erratic Diurnal Eating Patterns in Humans that Can Be Modulated for Health Benefits

Graphical Abstract



Highlights

- The daily eating pattern in healthy adults is highly variable from day to day
- More than half of the adults eat for 15 hr or longer every day
- Sleep duration parallels the fasting duration
- Reducing the daily eating duration can contribute to weight loss

Authors

Shubhroz Gill, Satchidananda Panda

Correspondence

satchin@salk.edu

In Brief

Gill et al. developed a smartphone app to monitor food intake in healthy adults, revealing frequent and erratic daily eating patterns rather than the self-reported three daily meals, leading to "metabolic jetlag." In a pilot study, overweight individuals using the app to eat for only 10–11 hr/day experienced sustained weight loss.

Gill & Panda, 2015, Cell Metabolism *22*, 1–10
November 3, 2015 ©2015 Elsevier Inc.
http://dx.doi.org/10.1016/j.cmet.2015.09.005

CellPress

Please cite this article in press as: Gill and Panda, A Smartphone App Reveals Erratic Diurnal Eating Patterns in Humans that Can Be Modulated for Health Benefits, Cell Metabolism (2015), http://dx.doi.org/10.1016/j.cmet.2015.09.005

Cell Metabolism
Clinical and Translational Report

A Smartphone App Reveals Erratic Diurnal Eating Patterns in Humans that Can Be Modulated for Health Benefits

Shubhroz Gill[1] and Satchidananda Panda[1,*]
[1]Regulatory Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA
*Correspondence: satchin@salk.edu
http://dx.doi.org/10.1016/j.cmet.2015.09.005

SUMMARY

A diurnal rhythm of eating-fasting promotes health, but the eating pattern of humans is rarely assessed. Using a mobile app, we monitored ingestion events in healthy adults with no shift-work for several days. Most subjects ate frequently and erratically throughout wakeful hours, and overnight fasting duration paralleled time in bed. There was a bias toward eating late, with an estimated <25% of calories being consumed before noon and >35% after 6 p.m. "Metabolic jetlag" resulting from weekday/weekend variation in eating pattern akin to travel across time zones was prevalent. The daily intake duration (95% interval) exceeded 14.75 hr for half of the cohort. When overweight individuals with >14 hr eating duration ate for only 10–11 hr daily for 16 weeks assisted by a data visualization (raster plot of dietary intake pattern, "feedogram") that we developed, they reduced body weight, reported being energetic, and improved sleep. Benefits persisted for a year.

INTRODUCTION

Life on earth has evolved in the context of a 24 hr periodicity in environmental conditions and a dependent daily rhythm in food availability and predator avoidance. Consequently, organisms have evolved endogenous circadian oscillators that allow them to anticipate and prepare for activity, sleep, and food intake at a specific time of the day. Both food ingestion and fasting can alter the metabolic state. Therefore, molecular responses to feeding and fasting exhibit temporal dynamics with a 24 hr period. The circadian oscillator and time of feeding act together to drive daily rhythms in gene expression and protein function such that the anticipation of and responses to feeding events are properly timed (Adamovich et al., 2014; Vollmers et al., 2009) every day. Genetic disruption of circadian rhythms in experimental animals and behavioral disruption of circadian rhythm among shift workers likely perturbs such temporal regulation and predisposes to metabolic diseases (Asher and Schibler, 2011). Frequent caloric intake in animal models of diet-induced obesity also dampens molecular circadian rhythms (Kohsaka et al., 2007). Conversely, recent studies in both nocturnal (mice) and diurnal (*Drosophila melanogaster*) model organisms have demonstrated that restricting the time of caloric intake to a window of 8–12 hr without altering the quantity or quality of diet can impart pleiotropic physiological benefits (Chaix et al., 2014; Gill et al., 2015; Hatori et al., 2012; Sherman et al., 2012; Zarrinpar et al., 2014). Such time-restricted feeding (TRF) supports a robust circadian rhythm and is associated with reduced adiposity, elevated lean mass, longer sleep duration, increased endurance, reduced systemic inflammation, decelerated cardiac aging, gut homeostasis, and improvement in other clinically relevant biomarkers.

Despite these observed benefits in model organisms, the applicability of TRF for human health has remained unknown, because the temporal aspect of human eating pattern is rarely measured. A lack of methods and parameters to describe the daily eating pattern in humans makes it difficult to ascertain whether eating events in humans are spread over a long enough segment of the 24 hr day such that there is an opportunity to reduce this duration. Furthermore, differences between metabolic properties of humans and those of small model organisms make it difficult to predict whether an 8–12 hr eating window will impart physiological benefits in humans. Nevertheless, accumulating data indicate that the temporal aspect of food intake, in addition to total caloric intake, can be an important determinant of predisposition to chronic diseases (Mattson et al., 2014), which now are the predominant cause of morbidity and mortality in developed nations (Bauer et al., 2014). Therefore, the objective longitudinal assessment of the content and the daily temporal pattern of human nutrition could have a major public health impact in terms of finding at-risk individuals for chronic disease and subsequent corrective action through lifestyle interventions.

Current methods to monitor human nutrition are subjective (such as questionnaires), remove subjects from their usual spatiotemporal niche (e.g., in-lab videography with limited food choices), or provide negative feedback that interferes with subjects' behavior (food diaries). Moreover, these methods gather information on the quantity and quality of nutrition and generally do not seek information regarding the time at which the food or beverage was consumed. The availability of smartphones presents an opportunity to objectively monitor human nutrition along with the advantage of complete control over feedback. The wide-ranging adoption of smartphones across age, gender, and socioeconomic segments can be leveraged to study behaviors of free-living individuals at scale in heterogeneous populations.

In this study, we developed and used a smartphone-based monitoring method with minimal feedback to collect the natural daily eating pattern of free-living healthy adults. In contrast to the conventional wisdom that modern humans eat 3 meals a day within a 12 hr interval, eating pattern was found to be largely erratic and differing between weekdays and weekends. In a pilot feasibility study, we tested whether reducing the eating interval to 10–11 hr without an overt attempt to change nutrition could lead to weight loss in healthy overweight individuals. Overall, the results presented here show that a large segment of the human adult population displays an erratic daily rhythm of eating-fasting, which can be manipulated to obtain desirable health benefits.

RESULTS AND DISCUSSION

A Smartphone App to Monitor Daily Eating Patterns

We built a smartphone software application (“app”) to longitudinally monitor the daily temporal pattern of caloric intake in free-living humans. To record an ingestive event, the participants used the camera function of the smartphone to take a picture of the food or beverage. Food pictures taken in the JPEG format were downscaled to 1/10 their original size on the device itself to reduce network data usage. Participants also had the option for textual entries to substitute for food/drink pictures when picture taking was difficult or when subjects forgot to log their picture entries. Immediately after data logging, the food picture and text entries along with timestamp and geolocation were immediately transferred to a server. Upon confirmation of successful data transfer, the data and the associated metadata (i.e., the timestamp and geolocation) were erased from the subject’s device, eliminating the possibility for the “feedback effect” of prior-recorded information upon present behavior.

We made use of push notifications as an orthogonal measurement technique for assessing the time of diet intake. These push notifications were manually triggered at a random time of the day, during the stated wakeful period of the subject and numbered 1–2 per day. The specific push notification query presented to the subject was dependent on the time when he/she responds to it, not when it was originally dispatched from the server. The push notification presented a query on the user’s device inquiring whether they ate/drank anything in the past 30 min. Subjects had to push a “Yes” or “No” button displayed on their screen, and their responses were recorded on the server. From such responses to push notifications sent at a random time during wakeful hours, we estimated the false negative rate (i.e., when the subject consumed food/beverage/water, but forgot to log the event) for our methodology to be 10.34%.

We monitored healthy, non-shift-worker adult males and females (Figures 1A and 1B, Table S1) for 3 weeks. After meeting the inclusion and exclusion criteria (Table S2) and signing an informed consent document during an office visit, subjects used the custom mobile application (Salk Metabolic App) installed on their smartphone to take pictures of every food, beverage, or water item they consumed, irrespective of volume or calorie, just prior to consumption (Figures S1 and S2). Appending a textual annotation describing the amount and the item(s) consumed to the pictures was optional.

The food pictures or occasional text entries (2.1% of all events) were further annotated by looking up the reference nutrition values from Calorie King or FNDDS (Six et al., 2011). Out of the 26,676 events recorded, 22% (5846) were water, 28% (7420) consisted of pre-packaged items with readily accessible nutrition information, and 50% (13,410) were mixed meals with multiple items. We hypothesized that the reported caloric intake should at least meet the resting energy expenditure or maintenance caloric (MC) intake (Roza and Shizgal, 1984). The average daily estimated caloric intake for the group (mean 1,947 kcal; 95% CI: 1,917–1,977) was more than their respective maintenance caloric intake (mean 1.233-fold over MC; 95% CI: 1.214–1.251). From push notifications, we had measured a false negative rate or underreporting of food/beverage/water to be 10.34%. Therefore, the actual caloric intake was likely little higher. The extra caloric intake likely accounts for activity above resting metabolism. There was no significant change in body weight during the 3-week reporting period (Table S1), indicating any potential feedback effect on weight loss due to recording food intake was absent.

Daily Eating Patterns in Humans

The timestamp of every ingestion event allowed analyses of the temporal aspect of eating. Aggregate data from 3 weeks of monitoring was used to assess eating pattern of the cohort. Caloric (i.e., >5 kcal) events populated a large segment of the 24 hr day (Figures 1C and 1D), and there were only 5 hr between 1–6 a.m. when the number of events/hr were <1% of total events (Figure 2A). The fraction of events with estimated energy content >5 kcal also reached its nadir (Figure 2B) in that interval. Because at the population level human digital activity reaches the trough between 2 and 4 a.m. (Golder and Macy, 2011), none of the subjects were self-reported shift-workers, and the reporting trough was close to 4 a.m. (Figures 2A and 2B), we considered 4 a.m. as the onset of “metabolic day” (i.e., events between 00:00:00 and 03:59:59 hr were included in the previous calendar day).

Surprisingly, in contrast to the self-reported 3 meals/day structure of meals from most of the participants, a breakfast-lunch-dinner temporal pattern was largely absent (Figure 2C). At the individual level, the number of events/day showed wide variation ranging from 4.22 ± 0.1 (mean + SEM) for the bottom decile to 15.52 ± 0.34 for the top decile, of which 3.33 ± 0.07 and 10.55 ± 0.24, respectively, were caloric events (Figure 2D).

From a subset of events that marked the beginning and end of an eating report (i.e., pictures of food at the start and of leftovers at the end of the meal), we calculated the average meal duration to be 14 min 36 s. Therefore, “events” from an individual with a <15 min inter-event interval were combined into one “meal.” At the group level, 25% of all meals were within 1 hr 25 min of another meal, and the median inter-meal interval was 3 hr 6 min. Only 25% of the meals occurred after >6 hr 41 min of fasting (Figure 2E).

The fraction of total calories consumed (starting at 4 a.m.) showed that less than 25% of caloric intake occurs before noon (Figure 2F). The percentages of total calories consumed after 6 p.m., 9 p.m., and 11 p.m. (and before 4 a.m. of the next day) were 37.5%, 12.2%, and 3.9%, respectively (Figure 2F). After adjusting for maintenance calories (MC) for each individual (Roza and Shizgal, 1984), the average cumulative percentage of

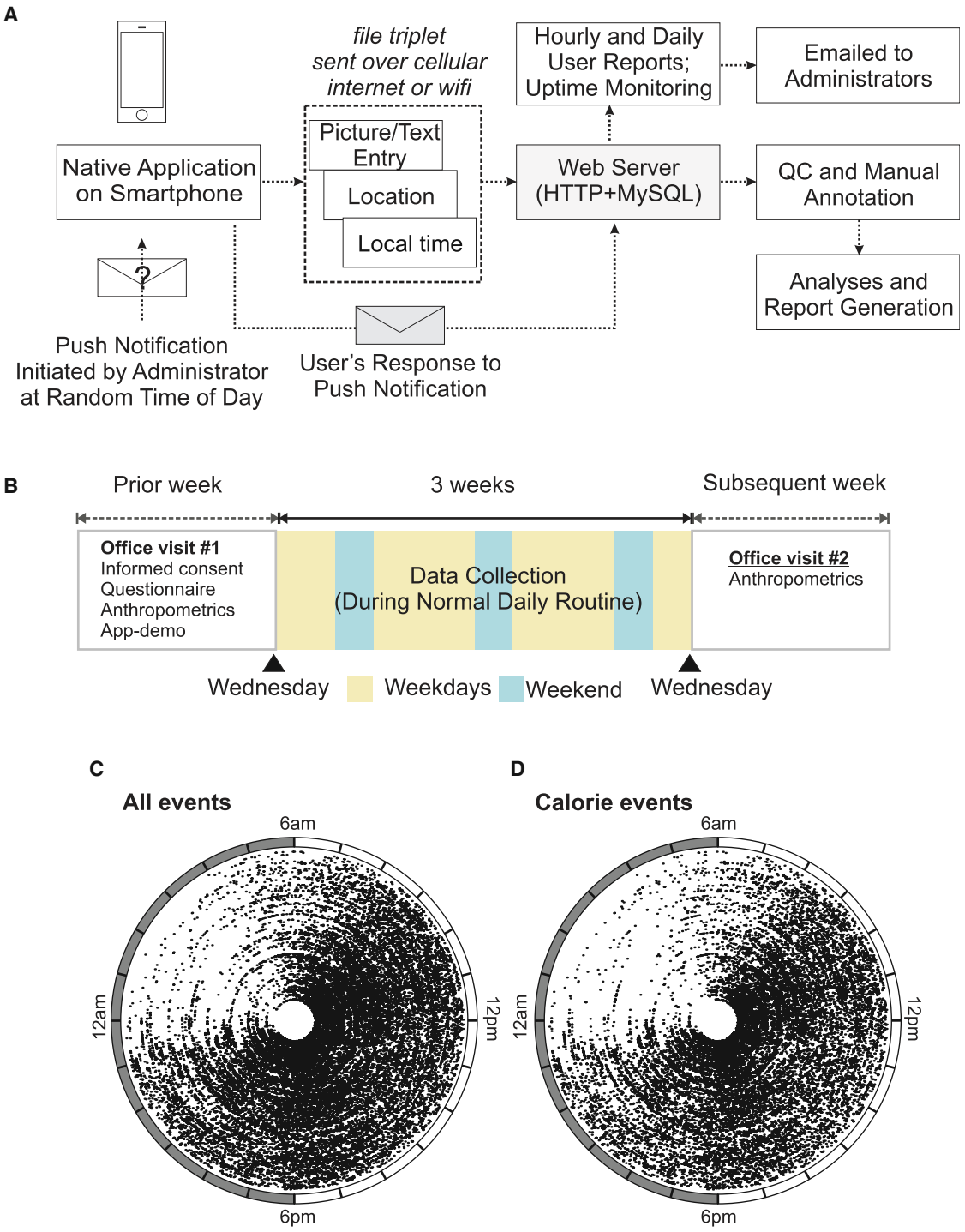


Figure 1. A Scalable Method to Monitor Daily Patterns of Dietary Intake in Humans

(A and B) Schematic (A) of the smartphone-based approach to monitor human eating pattern used to monitor all ingestion events for a 3-week period in healthy adults (B).

(C and D) Polar plot of all (C) or calorie-containing ($\geq$5 kcal) (D) ingestion events of each individual plotted against the time of day (radial axis) in each concentric circle. Data from 156 individuals are shown.

MC ingested over diurnal time showed the average time by which 50%, 70%, 90%, and 100% of MC were consumed to be 3:32 p.m., 5:04 p.m., 6:11 p.m., and 6:36 p.m., respectively (see Figure 2G for median values). In summary, there is a systematic bias toward consuming a larger portion of the daily caloric intake toward the late afternoon and evening hours. At the cohort

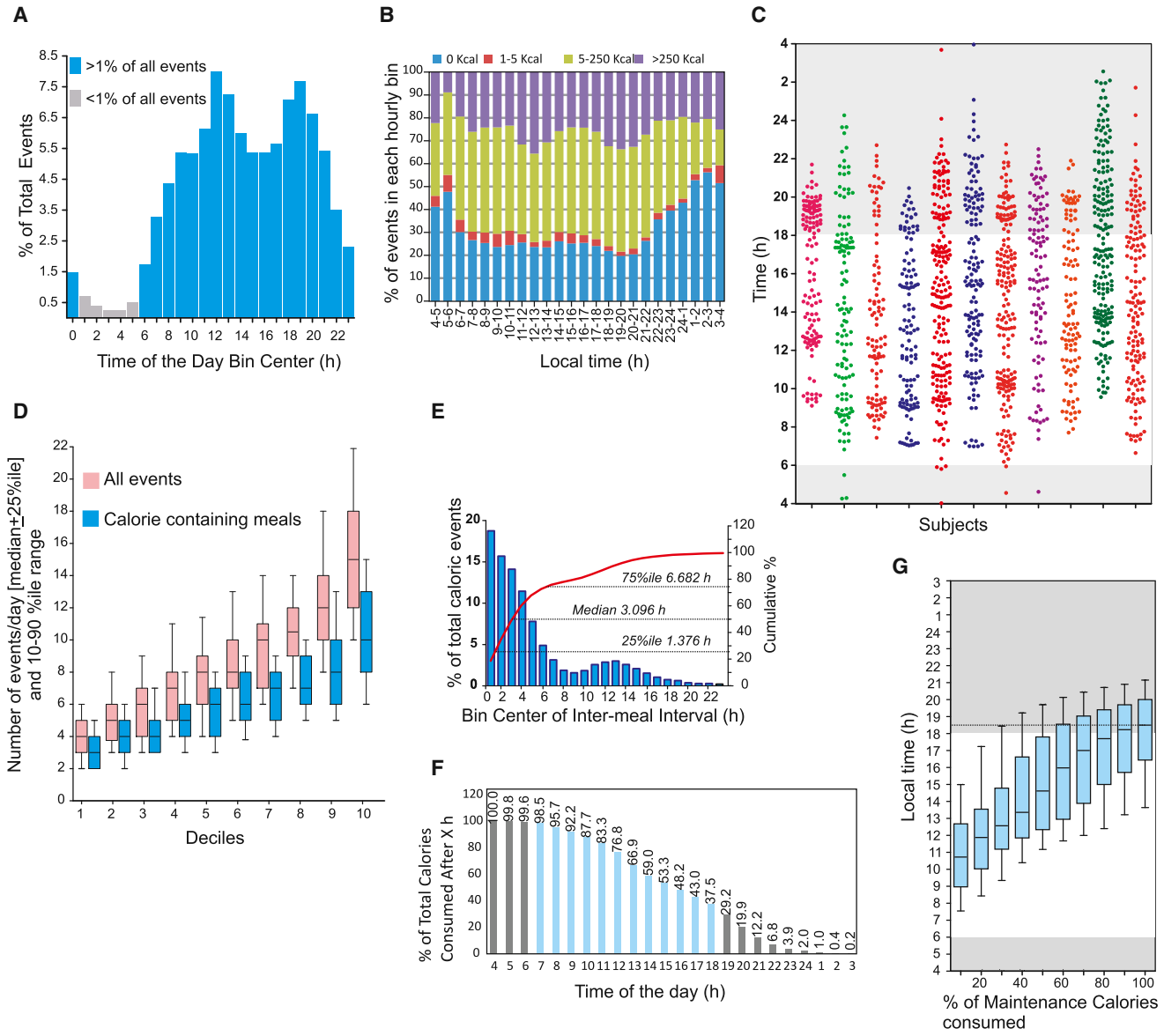


Figure 2. Human Eating Lacks 3-Meals-a-Day Structure

(A) Percentage of all ingestion events in 1 hr bin shows the nadir at 4 a.m.

(B) The fraction of events with <5 kcal also reaches its peak at 4 a.m. Therefore, we considered 4 a.m. (instead of midnight) as the beginning of the metabolic day.

(C) Representative scatter plot of ingestion events of 11 subjects during the observation period shows the lack of clustering of events into three principal bins for most individuals and a large variation in the total number of events.

(D) Number of ingestion events/day in all subjects binned into 10 deciles shows a wide distribution of number of total and calorie-containing events every day.

(E) Frequency distribution and cumulative frequency of inter-meal interval for the entire cohort.

(F) Percentage of calories remaining to be consumed in each hourly bin shows that >75% of all calories are consumed after midday.

(G) Time (median + 25%–50% range and 10%–90% interval) at which percentage of maintenance calories are consumed in 10% increments is shown.

level, in general, food consumed after 6:36 p.m. exceeded the maintenance calories requirement.

Eating Pattern Relative to Wakeful Hours

To compare the eating pattern with diurnal activity period, wrist actigraphy (Blood et al., 1997) data were collected from 47 randomly selected participants using a CamNTech Motionwatch 8, a device that measured both activity and light. A nighttime drop in activity and an absence of light was scored as time in bed. Integration of activity, light, and ingestion events allowed analyses of eating time relative to activity period (Figure 3A). The average time of activity onset and of time to bed showed large variation even among this non-shift worker cohort (Figure 3B). The median time interval between daily activity onset and first caloric intake was 1 hr 18 min, while the median time between the last caloric intake and going to bed was 2 hr 22 min

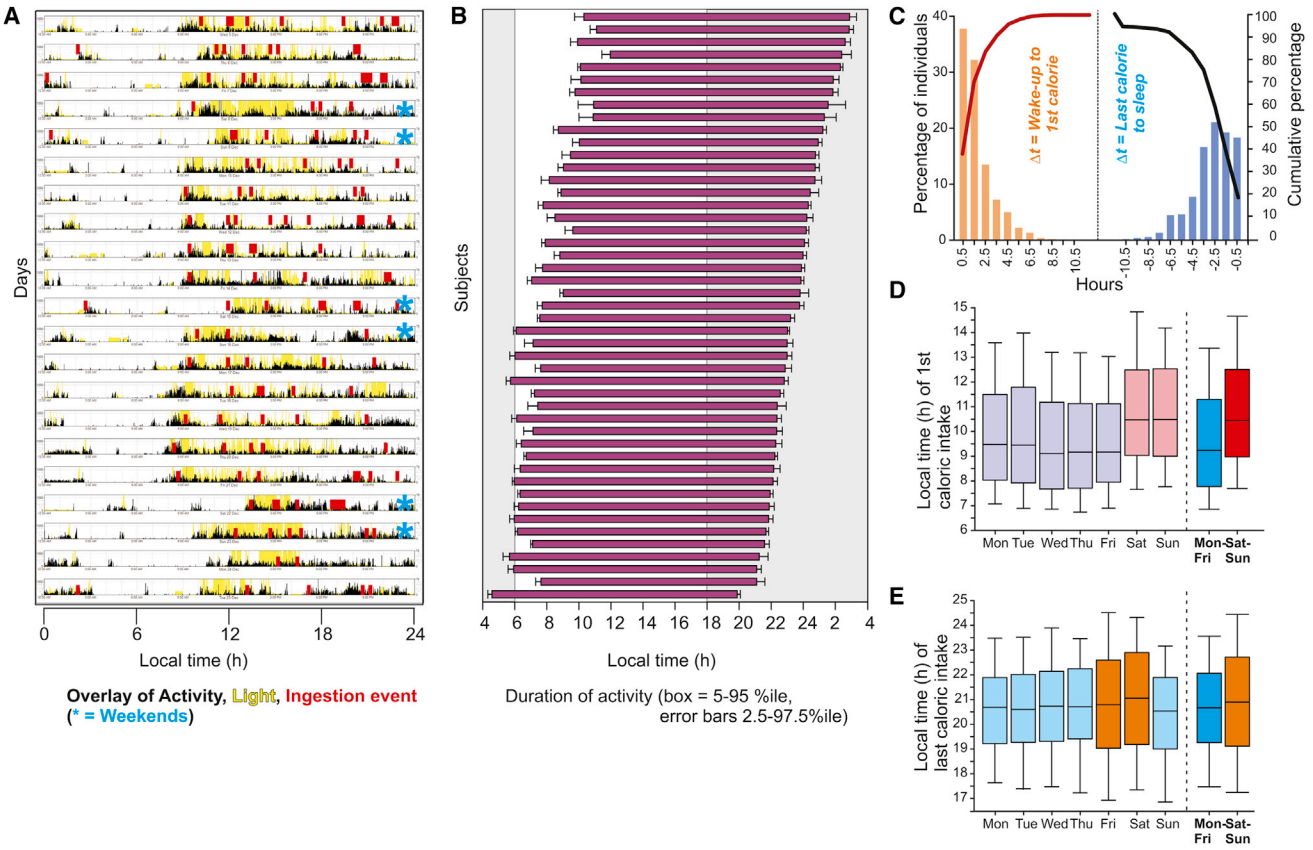


Figure 3. Activity and Eating Duration in Adult Humans

(A) Representative actogram and light exposure pattern (from a wrist-worn device) of a subject for 3 weeks overlaid with ingestion events (from smartphone app) show that the latter occur erratically throughout the active period.

(B) Wakeful activity duration in a subset of the subjects is shown. Each horizontal bar shows the interval between average wake up and bedtime (+SEM, up to 21 days of monitoring).

(C) Time interval between waking up and the first caloric ingestion or the last caloric ingestion and going to bed. Bars (orange and blue, y axis) indicate the percent of the individuals for whom actigraphy was performed with the indicated number of hours (x axis, 1 hr bins) from waking up to the first caloric intake or from the last caloric intake to sleep. Cumulative percentages (secondary y axis) are shown in color-matched lines.

(D and E) Median time of first (D) and last (E) caloric event of all individuals on different days of the week. Median (25%–75% interval in box, 10%–90% interval in lines) local time is shown.

(Figure 3C). Therefore, the total overnight fasting duration paralleled the time of inactivity (sleep) at night.

The day-to-day variation in the time of first or last caloric intake was spread over a few hours (Figure S3). Feeding after several hours of fasting is known to affect neuroendocrine metabolic pathways and adjust the phase of the circadian clock in peripheral organs (Vollmers et al., 2009) so that physiological state transitions from the fasting to the fed state. Changes in the sleep-wake cycle between social/work days and free/weekend days is similar to the circadian desynchrony arising from jet travel between time zones and is called social jetlag (Roenneberg et al., 2012). By analogy, we postulated that the variation in breakfast time between working/week days and free/weekend days likely affects the peripheral clocks in metabolic organs, causing metabolic desynchrony or "metabolic jetlag." The median breakfast (i.e., the first caloric intake event) time for the entire population changed on Saturday and Sunday (Figures 3D and S3), so we considered those two days as the "metabolic weekend." The time of last caloric intake did not significantly change in any of the days, but the variability was relatively large on Friday and Saturday (Figures 3E and S3). At the population level, the mean times of first caloric intake during weekdays and weekend were 9:21 a.m. (95% CI: 9:15–9:27 a.m.) and 10:26 a.m. (95% CI:10:15–10:37 a.m.), respectively. Delaying breakfast was more common than advancing it, with 40% of the subjects delaying breakfast by 1 hr or longer and 25% by 2.18 hr, while only 7% advanced their breakfast time by >1 hr. The time of last caloric intake was more variable than breakfast. The average last caloric intake time was advanced by >1 hr in 17%, while only 15% delayed the time of last caloric intake by >1 hr on the weekend.

Eating Duration and Eating Pattern

Having observed a large variance in the first and last caloric intake (Figure S3) and the absence of a clear 3 meals/day eating pattern (Figures 2C and 2D) for most subjects, a potentially better description of an individual's eating pattern could be the daily duration of caloric intake. Because food intake triggers postprandial changes in neuroendocrine state that can take minutes

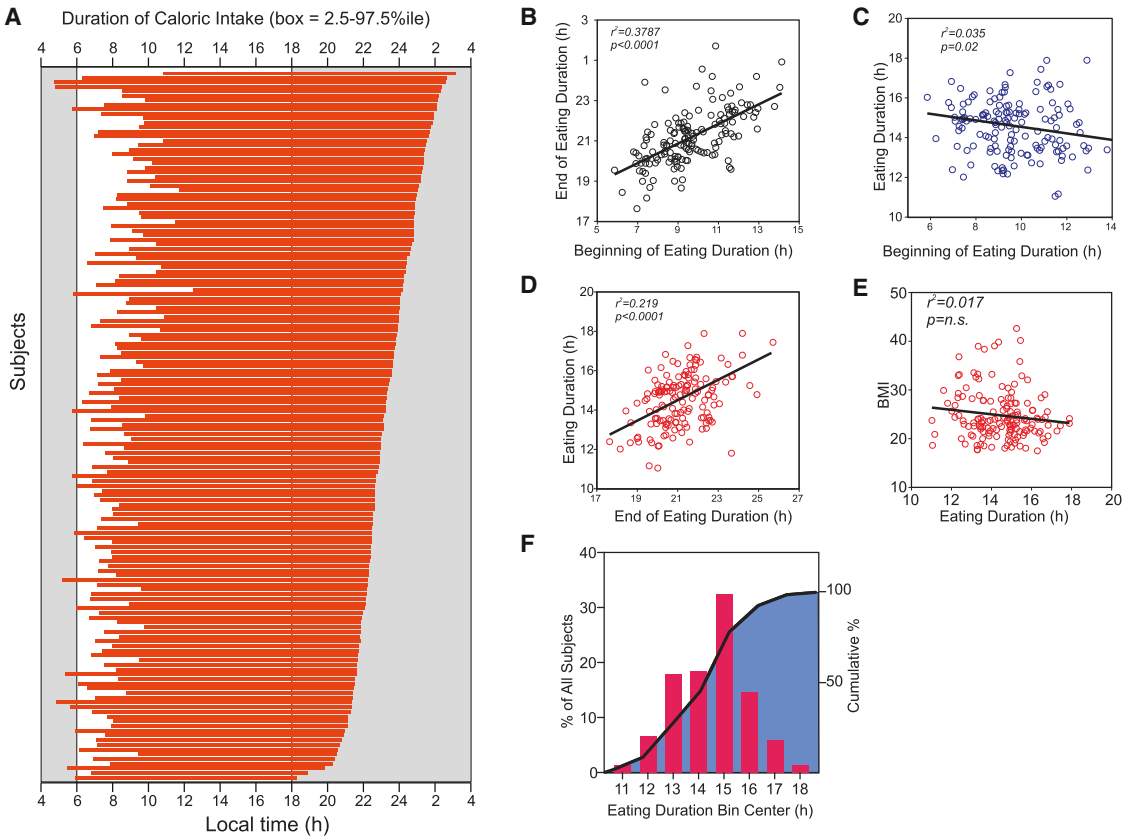


Figure 4. Daily Duration of Caloric Intake

(A) Eating duration of all individuals is shown in the order of late (top) to early (bottom) nighttime fasting onset time.

(B) The time of last caloric intake weakly positively correlates with the time of first caloric intake.

(C and D) The daily duration of eating does not correlate with the time of first caloric intake (C) but weakly positively correlates with the time of the last caloric intake (D).

(E and F) The subjects' daily eating duration does not correlate with their BMI (kg/m^2) (E), frequency distribution (red bars), and cumulative percentage (black line and blue filled area) of eating duration (hours) (F).

to hours to return to resting or fasting state, eating too often could clamp the physiological state between frequent meals to the post-prandial state. Therefore, we defined the daily eating duration as the time interval (4 a.m. onward) that contained 95% (2.5–97.5 percentile) of all intake events during the monitoring period (Figure 4A). This approach for arriving at the eating duration from aggregate data over several days is tolerant of occasional non-reporting of some random eating events. Breakfast time weakly positively correlated with the last caloric intake (r^2 = 0.379), so that individuals with earlier breakfast also had their last caloric intake earlier in the evening (Figure 4B). The eating duration better correlated with the time of last caloric intake (r^2 = 0.215) than with the time of breakfast (r^2 = 0.035) (Figures 4C and 4D) or with BMI (r^2 = 0.017) (Figure 4E). The median daily eating duration was 14 hr 45 min, and only 9.7% of the subjects had a daily eating duration <12 hr (Figure 4F) long. The weak correlation (r^2 = 0.017) between the eating duration and BMI could be due to the limited sample size, the heterogeneity of the participants in terms of gender and age, and the fact that the eating pattern recorded in the monitoring period is a short-term snapshot of a person's long-term diet-related behaviors.

Restricting Eating Duration Reduces Body Weight

Many factors, including nutrition quality, quantity, physical activity, and genetics contribute to obesity. Although we did not find a simple correlation between BMI and eating duration, we wanted to test whether longer eating duration and erratic eating pattern are contributing factors in subjects with co-occurrence of >25 BMI and >14 hr eating duration. We tested if reducing the eating duration and metabolic jetlag associated with weekday/weekend differences in a subset of individuals would lead to reduction in body weight. We recruited 8 individuals with >14 hr eating duration for a 16-week pilot intervention study, such that each individual's own baseline data served as the control (Figure 5A) condition. Individualized "feedogram" graphics representing a temporal raster plot of ingestion events in successive days were constructed (Figure 5B). After obtaining informed consent for intervention, the participants were provided data regarding their eating pattern accrued in the 3-week baseline period (95% eating duration, variance in first and last caloric intake, and weekday-weekend metabolic jetlag) and were shown their own baseline "feedogram" prior to the intervention. Because in rodents a daily eating period of up to 12 hr improves metabolic fitness (Chaix et al., 2014),

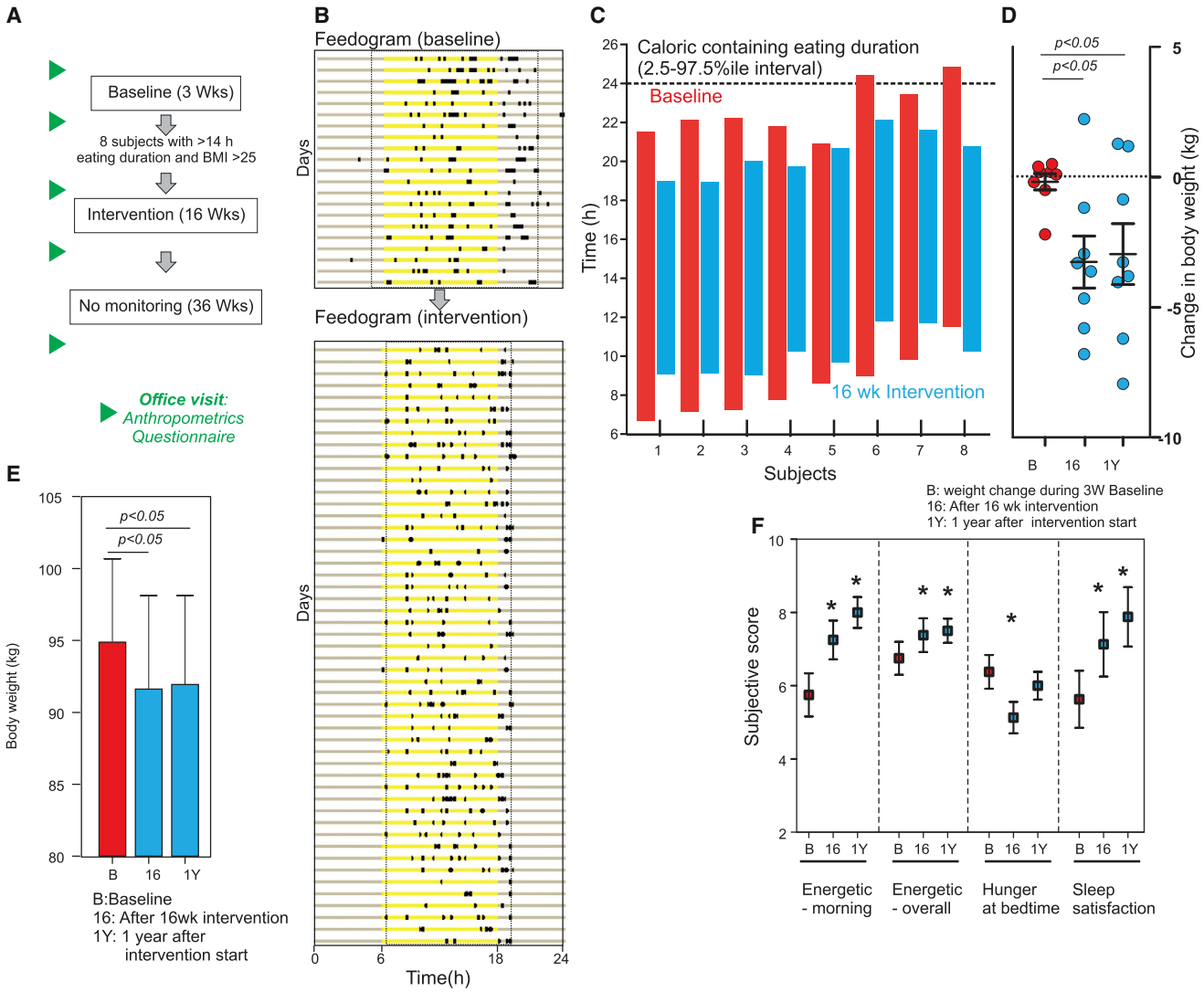


Figure 5. Improved Eating Pattern Reduces Body Weight in Healthy Overweight Individuals

(A) Schematic of the study design to test the effect of eating pattern on body weight.

(B) Representative "feedogram" of a participant during baseline and during intervention. The times of ingestion events are denoted as prominent black rectangles along the 24 hr day represented in each horizontal line (x axis). Yellow represents the time between 6 a.m. and 6 p.m. Eating duration during baseline and intervention is shown as broken lines.

(C) The daily eating duration of each individual during baseline (red) and intervention (blue) plotted against the local time (y axis).

(D) Scatter plot and average (±SEM) change in body weight in 8 participants during 3 weeks of baseline monitoring, after 16 weeks of intervention and after 1 year.

(E) Average (+SEM) body weight at the end of 3 weeks of baseline, after 16 weeks of monitored intervention and at 1 year.

(F) Average (+SEM) of subjective measures of energy level, hunger, and sleep in subjects. These subjective measures were assessed on a scale of 1–10, with 10 being the preferred (healthier) end of the range. Higher numbers thus indicated healthier values for the quantity, i.e., more energetic, less hunger at bedtime, and more sleep satisfaction. *$p < 0.05$, t test.

the participants were requested to reduce their caloric-containing eating duration to a self-selected window of 10–12 hr and to consistently follow this duration during both weekdays and weekends so that the metabolic jetlag could be minimized. No overt suggestion concerning nutrition quality, quantity, or caloric content was provided. The individuals continued logging their food pictures using the same app as used in the baseline period for the next 16 weeks and also received a weekly summary of their feedograms and daily eating duration.

All subjects reduced their eating duration (average reduction: 4 hr 35 min; 95% CI: 3 hr 30 min–5 hr 40 min; $p < 0.001$), and their weekday/weekend metabolic jetlag was also reduced to <1 hr (Figures 5B and 5C). The participants showed a reduction in total body weight (average loss 3.27 kg; 95% CI: 0.9081–5.624 kg) and, accordingly, excess body weight (Figures 5D and 5E, Table S3) and BMI (average reduction 1.15 kg/m^2; 95% CI: 0.3247–1.980 kg/m^2). In a subjective self-assessment of sleep satisfaction, hunger at bedtime, and energy level

(in the mornings, and overall over the past few days), statistically significant improvement was observed (Figure 5F). All participants voluntarily expressed an interest in continuing unsupervised with the 10–11 hr time-restricted eating regimen after the conclusion of the 16-week supervised intervention. After 36 weeks (1 year since the intervention began), the participants maintained weight loss and sleep improvement and felt more energetic (Figures 5D–5F, Table S3).

Although the participants were not overtly asked to change nutrition quality or quantity, reducing the eating duration led to reduced estimated caloric intake. Unlike mice, where reducing the eating duration to ~10 hr does not alter total caloric intake (Hatori et al., 2012), our human intervention cohort reduced the estimated daily caloric intake (average reduction 20.26%; 95% CI 4.92%–35.6%; paired t test $p < 0.05$). Humans consume heterogeneous food types in a time-of-the-day-dependent manner (Figure S4), e.g., coffee is almost always consumed in mornings, while alcohol at night. As a result, during the intervention, it was not the case that items that would have otherwise (in the baseline period) been consumed in the designated 14 hr nighttime fasting hours had been moved to the self-selected 10 hr feeding period during the intervention. Instead, the person would simply not consume such an item rather than consume it at the wrong time of day. This could be one potential explanation for the reduction in caloric intake.

Discussion

Collecting human nutrition information in the free-living condition has been a persistent challenge. Recording dietary intake using text entries, selecting from a large library of food items, and specifying the portion size is a ubiquitous feature found in most nutrition apps. Although such apps improve adherence relative to the traditional diary log (Carter et al., 2013), data logging can be cumbersome for mixed meals, and consequently, users may not bother to log small snacks. Furthermore, portion size reporting can be subjective. By adopting an approach centered on food pictures with an optional user-side annotation together with infrequent, but randomly timed, push notifications, we reduced the barrier to data recording. Server side annotation of the picture metadata ensured uniformity across the cohort. Supervised crowd-sourced annotation can make this approach scalable to large cohorts.

By overlaying the daily patterns of food and beverage intake, activity-rest, and light exposure, we could uncover relationships among them (Figures 3A–3C). Data integration from multiple such longitudinal data streams has immense promise for disease prognosis. Although the subjects did not have any chronic medical conditions, they also logged their consumption of vitamins, supplements, and occasional over-the-counter medications for minor ailments, thus offering a temporal pattern of drug and supplement use (Figure S4). Greater than 50% of the mammalian transcriptome exhibits diurnal rhythms in a tissue-specific manner (Zhang et al., 2014), the gut microbiome shows daily rhythms (Thaiss et al., 2014), the timing of food affects these rhythms in peripheral organs (Vollmers et al., 2009), and the targets of a large number of FDA approved drugs show circadian expression (Zhang et al., 2014). Therefore, monitoring the timing of drug intake relative to the sleep-wake or feeding-fasting cycle can have a significant impact on disease prognosis and unraveling interactions among food, sleep, and drugs in free-living individuals.

Formally, our work introduces a method and the critical defining parameters for describing the diurnal and longer-term temporal characteristics of nutrition in humans. By creating a scalable method to longitudinally monitor human nutrition in an evidence-based manner, we discovered that the daily eating pattern even among healthy non-shift young workers is highly variable from day to day. For more than half of the participants in the baseline monitoring study, the eating pattern is erratic, and energy intake events span over a large fraction of a 24 hr day, with a relatively short fasting period (Figure 4). Although the first caloric intake after leaving the bed happened within 1 hr 18 min (median value) (Figure 3C), less than 25% of the daily caloric intake occurred before noon, while 37.5% was consumed after 6 p.m. (Figure 2F). This suggests that breakfast is relatively small in terms of energetic input and that major caloric intake is delayed until later in the afternoon or evening in this set of relatively young (Table S1) non-shift-worker subjects. To address the universality of our observations concerning eating patterns, this method may be extended to a larger population spread over different geographical regions, work schedules (e.g., shift-workers, retired individuals, nurses, pilots), age groups, and/or cultures. It would also be useful to describe the diurnal patterns of caloric intake in humans that do not have a modern lifestyle influenced by electricity, such as hunter-gather societies.

Individuals in our study largely ate throughout the wakeful hours (Figure 3). Consequently, sleep duration and quality largely dictated the eating pattern. Furthermore, since the sleep pattern changes between weekdays (workdays) and weekends, leading to social jetlag (Roenneberg et al., 2012; Wittmann et al., 2006), the breakfast time also changes between weekdays and weekends. These changes in breakfast time are analogous to a person traveling across time zones every weekend and can be described as metabolic jetlag. The intricate connection we observed between sleep and overnight fasting duration suggests that the observed relationship between a short sleep duration and predisposition to metabolic diseases (Cappuccio et al., 2011; Copinschi et al., 2014) may be partly explained by the reduced duration of overnight fasting. Similarly, the reported correlation between social jetlag and BMI may also involve metabolic jetlag. The increased daily eating duration likely contributes to increased caloric intake. A change in eating pattern between days (e.g., weekday versus weekend) can affect time-of-day/night-specific changes in food intake from specific food groups (Figure S4). Therefore, one mode by which reduced sleep duration contributes to the increased risk for metabolic diseases could be the increased daily eating duration and associated changes in caloric intake and nutrition quality.

We did not find a simple positive correlation between the daily eating duration and BMI in our cohort. This may be for several reasons, including a limited sample size, heterogeneity of the cohort, and a likely scenario that individuals with long eating duration may also have more physical activity. Nevertheless, reducing the temporal eating period in a feasibility study imparted measurable benefits of clinically relevant magnitude in terms of body weight reduction and sleep improvement without increasing the subjective sense of hunger. This relatively large

effect on body weight reduction, even in the small intervention cohort, implies that the benefits might result from multiple changes: restoration of the diurnal rhythm of feeding/fasting, reduction of the weekday/weekend metabolic jetlag, and a reduction in the daily caloric intake. Some benefits of TRF might arise from caloric reduction (CR). At the same time, we cannot rule out the possibility that some benefits of CR in vertebrates including humans might be from TRF, as most CR studies involve caloric intake within a defined time frame. Nevertheless, if time restriction under free-living condition inadvertently leads to caloric reduction, TRF as a method to reduce caloric intake is a more attractive option, as individuals, caregivers, case managers, physicians, and scientists do not have to adopt expensive and laborious methods to accurately track caloric count. Hence, irrespective of mechanism, time restriction offers an effective approach to improve health.

While the relative contribution of daily eating pattern, calories, and nutrition quality to multifaceted health improvement in humans should be examined in detail in future studies, our results highlight that suitable manipulation of the diurnal temporal pattern of caloric intake is a feasible therapeutic approach for improving human health in the free-living condition, in spite of the vast variety of food and beverage types consumed by the average person from day to day. This opens up the possibility for utilizing this strategy by itself or in combination with existing approaches for health improvement.

EXPERIMENTAL PROCEDURES

This study was approved by the IRB of the Salk Institute. Participants were recruited during 2012–2013 through a newspaper advertisement, paper flyers, and online advertisements. Inclusion and exclusion criteria were determined by an online questionnaire and an in-person interview. All subjects provided written informed consent during the first office visit and were asked to record all of their food, beverage, and water intake using the smartphone app. Subjects' height and weight were measured using a calibrated scale and tape measure at the beginning and end of the 3-week baseline period. Participants were nominally compensated for their time and effort.

For the baseline study aimed at observing eating patterns in free-living adults, data were collected from a Tuesday/Wednesday midnight to another Tuesday/Wednesday midnight 21 days later. Subjects received instructions to record every item consumed (food, drink, water) regardless of its size using the app on their smartphone. The leftovers from items that were not completely consumed were to be recorded again, but described as such in the annotation. Days with fewer than 3 total events (including non-caloric content items) were flagged and verified with the participants for any observed fasting day or whether they forgot to log data.

Subjects from the baseline study with BMI > 25 kg/m^2 and whose intake interval exceeded 14 hr were offered the opportunity to participate in the intervention study. After a detailed presentation on the known benefits of time-restricted feeding in rodents, data on their own eating pattern, and what the intervention study would entail, subjects chose a 10 hr eating interval of their choice and were to limit all non-water intake (including coffee and tea) to that 10 hr interval. Eight subjects entered the study: 5 males (age 34.4 ± 2.9 years, weight 96.7 ± 4.8 kg, BMI 31.77 ± 2.05 kg/m^2), 3 females (age 36.3 ± 4.3 years, weight 91.8 ± 15 kg, BMI 34.91 ± 3.84 kg/m^2; average ± SEM). Ethnicity: 2 Hispanics, 6 non-Hispanic or Latino. Race: 3 Asian, 4 white, 1 more than one race. To identify undiagnosed fasting hypoglycemia, a fasting blood metabolic panel was performed at a certified clinic prior to the intervention study for every subject. During the 16-week-long intervention study, the participants were instructed to log in to a personalized website every week to check their eating duration. Anthropometrics were performed prior to the start of the intervention, just after the completion of the 16-week-long intervention and 1 year after the start of the intervention. As the goal for this feasibility pilot study was to test TRF in healthy overweight subjects with no recent history of fasting blood sugar, cholesterol, and triglycerides outside the reference range, we did not thoroughly investigate clinical biomarkers of metabolic diseases.

Software

The smartphone application was coded in Objective-C. When the application is run for the first time, it generates a unique 20-character alphanumeric code to identify the device as well as records the user's Push Notification Service identifier in a database. Access to the application was restricted to study participants by a one-time username-based validation procedure. Food pictures taken in JPEG format were downscaled to 1/10 their original size on the device itself to reduce data usage. The functionality of each app tab is shown in Figure S1. Push notifications were manually triggered at a random time of the day during the stated wakeful period of the subject and numbered 1–2 per day. The specific push notification-based query presented to the subject is dependent on the time when he/she sees it, not when it was originally initiated from the server. The presented query and the response were recorded on the server.

The Salk Metabolic Study app was available from the Apple App Store during the study period. To limit usage to subjects who provided a written informed consent, the app activation required a randomly generated unique activation key for each participant. Recently, the app was revised, upgraded, and re-written to run on the latest operating systems in iOS and Android devices under the name "myCircadianClock." Under Salk IRB approved study, the app is available to adults living in the U.S. Potential users can visit the website (http://mycircadianclock.mycircadianclock.org/) to review the ongoing study objectives and the consent document. If they consent to participate through electronic consenting, they are sent an activation key to use the app.

Data Annotation and Analysis

Duplicate pictures were first automatically removed by comparing MD5 checksums, which serve as a fingerprint of a file. A second round of de-duplication was performed by manually inspecting the contents of sequentially received food/drink pictures. De-duplicated data were then independently annotated by two researchers for multiple characteristics that describe each item. The items and estimated portion size were also looked up in FNDDS or CalorieKing website for estimated caloric content. A third researcher tallied the annotations and where discrepancies existed; all three individuals conferred and arrived at a consensus description. In recognition of the fact that lifestyles and work schedules are controlled more by conventional time than solar time, the time data shown herein referred to as "local time" incorporate daylight savings time when applicable. In other words, depending on whether daylight savings time was in effect (March–November) or not (November–March), a local time of 9.5 would refer to 9:30 a.m. PDT or 9:30 a.m. PST for an event recorded in California. Polar plots were generated using Mathematica 9 (Wolfram). All other plots and statistical analyses were prepared using Prism 5 (GraphPad Software).

Feedogram

To generate a feedogram raster plot, events were binned into 15 min intervals starting at midnight. Corresponding to each such bin, if an event was recorded in it, the segment in the raster plot was colored black. If no event was present, the segment was colored gray if it was between 6 p.m. and 6 a.m. and yellow otherwise.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cmet.2015.09.005.

AUTHOR CONTRIBUTIONS

Both S.G. and S.P. conceived the idea, designed the experiments, analyzed data, and wrote the manuscript. S.G. wrote the smartphone application and server-side software, recruited subjects, and collected data.

CONFLICTS OF INTEREST

S.G. and S.P. disclose financial interest in Cadius Health, a lifestyle modification company. Cadius Health did not support the aforementioned work, and at present this company has no right to any technology or intellectual property developed as part of this work.

ACKNOWLEDGMENTS

We thank our research assistants Sami Michishita and Toriana Dabkowski for invaluable help with the data collection and annotation. This work was partially supported by NIH grants DK091618, EY016807, and DK063491, AFAR grant #M14322, Leona M. and Harry B. Helmsley Charitable Trust's grant #2012-PG-MED002, and Joe W. and Dorothy Dorsett Brown Foundation. S.G. was supported by H.A. and Mary K. Chapman Trust and Aginsky Research Scholar Award. Additional support came from the Glenn Center for Aging and Salk Innovation Grant.

Received: April 27, 2015
Revised: July 20, 2015
Accepted: September 8, 2015
Published: September 24, 2015

REFERENCES

Adamovich, Y., Rousso-Noori, L., Zwighaft, Z., Neufeld-Cohen, A., Golik, M., Kraut-Cohen, J., Wang, M., Han, X., and Asher, G. (2014). Circadian clocks and feeding time regulate the oscillations and levels of hepatic triglycerides. Cell Metab. *19*, 319–330.

Asher, G., and Schibler, U. (2011). Crosstalk between components of circadian and metabolic cycles in mammals. Cell Metab. *13*, 125–137.

Bauer, U.E., Briss, P.A., Goodman, R.A., and Bowman, B.A. (2014). Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. Lancet *384*, 45–52.

Blood, M.L., Sack, R.L., Percy, D.C., and Pen, J.C. (1997). A comparison of sleep detection by wrist actigraphy, behavioral response, and polysomnography. Sleep *20*, 388–395.

Cappuccio, F.P., Cooper, D., D'Elia, L., Strazzullo, P., and Miller, M.A. (2011). Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. Eur. Heart J. *32*, 1484–1492.

Carter, M.C., Burley, V.J., Nykjaer, C., and Cade, J.E. (2013). Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J. Med. Internet Res. *15*, e32.

Chaix, A., Zarrinpar, A., Miu, P., and Panda, S. (2014). Time-restricted feeding is a preventative and therapeutic intervention against diverse nutritional challenges. Cell Metab. *20*, 991–1005.

Copinschi, G., Leproult, R., and Spiegel, K. (2014). The important role of sleep in metabolism. Front. Horm. Res. *42*, 59–72.

Gill, S., Le, H.D., Melkani, G.C., and Panda, S. (2015). Time-restricted feeding attenuates age-related cardiac decline in Drosophila. Science *347*, 1265–1269.

Golder, S.A., and Macy, M.W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science *333*, 1878–1881.

Hatori, M., Vollmers, C., Zarrinpar, A., DiTacchio, L., Bushong, E.A., Gill, S., Leblanc, M., Chaix, A., Joens, M., Fitzpatrick, J.A., et al. (2012). Time-restricted feeding without reducing caloric intake prevents metabolic diseases in mice fed a high-fat diet. Cell Metab. *15*, 848–860.

Kohsaka, A., Laposky, A.D., Ramsey, K.M., Estrada, C., Joshu, C., Kobayashi, Y., Turek, F.W., and Bass, J. (2007). High-fat diet disrupts behavioral and molecular circadian rhythms in mice. Cell Metab. *6*, 414–421.

Mattson, M.P., Allison, D.B., Fontana, L., Harvie, M., Longo, V.D., Malaisse, W.J., Mosley, M., Notterpek, L., Ravussin, E., Scheer, F.A., et al. (2014). Meal frequency and timing in health and disease. Proc. Natl. Acad. Sci. USA *111*, 16647–16653.

Roenneberg, T., Allebrandt, K.V., Merrow, M., and Vetter, C. (2012). Social jetlag and obesity. Curr. Biol. *22*, 939–943.

Roza, A.M., and Shizgal, H.M. (1984). The Harris Benedict equation reevaluated: resting energy requirements and the body cell mass. Am. J. Clin. Nutr. *40*, 168–182.

Sherman, H., Genzer, Y., Cohen, R., Chapnik, N., Madar, Z., and Froy, O. (2012). Timed high-fat diet resets circadian metabolism and prevents obesity. FASEB J. *26*, 3493–3502.

Six, B.L., Schap, T.E., Kerr, D.A., and Boushey, C.J. (2011). Evaluation of the Food and Nutrient Database for Dietary Studies for use with a mobile telephone food record. Journal of food composition and analysis *24*, 1160–1167.

Thaiss, C.A., Zeevi, D., Levy, M., Zilberman-Schapira, G., Suez, J., Tengeler, A.C., Abramson, L., Katz, M.N., Korem, T., Zmora, N., et al. (2014). Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. Cell *159*, 514–529.

Vollmers, C., Gill, S., DiTacchio, L., Pulivarthy, S.R., Le, H.D., and Panda, S. (2009). Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. Proc. Natl. Acad. Sci. USA *106*, 21453–21458.

Wittmann, M., Dinich, J., Merrow, M., and Roenneberg, T. (2006). Social jetlag: misalignment of biological and social time. Chronobiol. Int. *23*, 497–509.

Zarrinpar, A., Chaix, A., Yooseph, S., and Panda, S. (2014). Diet and feeding pattern affect the diurnal dynamics of the gut microbiome. Cell Metab. *20*, 1006–1017.

Zhang, R., Lahens, N.F., Ballance, H.I., Hughes, M.E., and Hogenesch, J.B. (2014). A circadian gene expression atlas in mammals: implications for biology and medicine. Proc. Natl. Acad. Sci. USA *111*, 16219–16224.

Development of the social brain in adolescence

Sarah-Jayne Blakemore

Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, UK
Email: s.blakemore@ucl.ac.uk

Summary

The brain has evolved to understand and interact with other people. We are increasingly learning more about the neurophysiological basis of social cognition and what is known as the social brain, that is the network of brain regions involved in understanding others. This paper focuses on how the social brain develops during adolescence. Adolescence is a time characterized by change – hormonally, physically, psychologically and socially. Yet until recently this period of life was neglected by cognitive neuroscience. In the past decade, research has shown that the brain develops both structurally and functionally during adolescence. Large-scale structural MRI studies have demonstrated development during adolescence in white matter and grey matter volumes in regions within the social brain. Activity in some of these regions, as measured using fMRI, also shows changes between adolescence and adulthood during social cognition tasks. I will also present evidence that theory of mind usage is still developing late in adolescence. Finally, I will speculate on potential implications of this research for society.

This is the first in a series of three articles on a new clinical concept called violence adapting syndrome

DECLARATIONS

Competing interests
None declared

Funding
None

Ethical approval
Not applicable

Guarantor
SJB

Contributorship
SJB is the sole contributor

Acknowledgements
The author is supported by a Royal Society University Research Fellowship

Reviewer
David Misselbrook

Humans are exquisitely social

Humans are an exquisitely social species. We are constantly reading each other's actions, gestures and faces in terms of underlying mental states and emotions, in an attempt to figure out what other people are thinking and feeling, and what they are about to do next. This is known as theory of mind or mentalising. Developmental psychology research on theory of mind has demonstrated that the ability to understand others' mental states develops over the first four or five years of life. While certain aspects of theory of mind are present in infancy,[1] it is not until around the age of four years that children begin explicitly to understand that someone else can hold a belief that differs from one's own, and which can be false.[2] An understanding of others' mental states plays a critical role in social interaction because it enables us to work out what other people want and what they are about to do next, and to modify our own behaviour accordingly.

The social brain

Over the past 15 years, a large number of independent neuroimaging studies have shown remarkable consistency in identifying the brain regions that are involved in theory of mind or mentalising. These studies have employed a wide range of stimuli including stories, sentences, words, cartoons and animations, each designed to elicit the attribution of mental states (see Amodio and Frith[3] for review). In each case, the mentalising task resulted in the activation of a network of regions including the posterior superior temporal sulcus (pSTS) at the temporo-parietal junction

Downloaded from jrs.sagepub.com at GEORGETOWN UNIV MED CTR on May 25, 2015

(TPJ), the temporal poles and the dorsal medial prefrontal cortex (mPFC; see Burnett and Blakemore[4]). The agreement between neuroimaging studies in this area is remarkable and the consistent localisation of activity within a network of regions including the pSTS/TPJ and mPFC, as well as the temporal poles, suggests that these regions are key to the process of mentalising.

Brain lesion studies have consistently demonstrated that the superior temporal lobes[5] and PFC[6] are involved in mentalising, as damage to these brain areas impairs mentalising abilities. Interestingly, one study reported a patient with extensive PFC damage whose mentalising abilities were intact,[7] suggesting that this region is not necessary for mentalising. However, there are other explanations for this surprising and intriguing finding. It is possible that, due to plasticity, this patient used a different neural strategy in mentalising tasks. Alternatively, it is possible that damage to this area at different ages has different consequences for mentalising abilities. The patient described by Bird and colleagues[7] had sustained her PFC lesion at a later age (62 years) than most previously reported patients who show impairments on mentalising tasks. Perhaps mPFC lesions later in life spare mentalising abilities, whereas damage that occurs earlier in life is detrimental. It is possible that mPFC is necessary for the acquisition of mentalising but not essential for later implementation of mentalising. Intriguingly, this is in line with recent data from developmental fMRI studies of mentalising, which suggest that the mPFC contributes differentially to mentalising at different ages.

Development of mentalising during adolescence

There is a rich literature on the development of social cognition in infancy and childhood, pointing to step-wise changes in social cognitive abilities during the first five years of life. However, there has been surprisingly little empirical research on social cognitive development beyond childhood. Only recently have studies focused on development of the social brain beyond early childhood, and these support evidence from social psychology that adolescence represents a period of significant social development. Most researchers in the field use the onset of puberty as the starting point for adolescence. The end of adolescence is harder to define and there are significant cultural variations. However, the end of the teenage years represents a working consensus in Western countries. Adolescence is characterized by psychological changes in terms of identity, self-consciousness and relationships with others. Compared with children, adolescents are more sociable, form more complex and hierarchical peer relationships, and are more sensitive to acceptance and rejection by peers.[8] Although the underlying factors of these social changes are most likely to be multifaceted, one possible cause is development of the social brain.

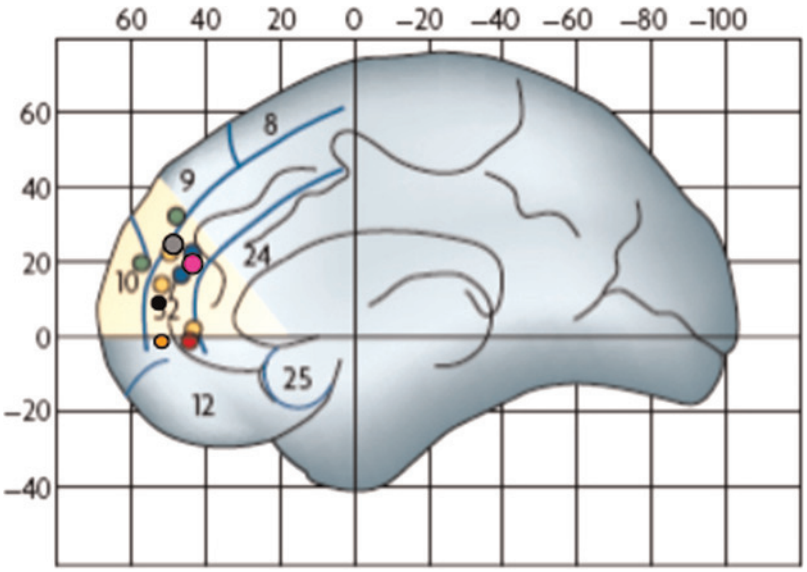
Recently, a number of fMRI studies have investigated the development during adolescence of the functional brain correlates of mentalising. These studies have used a wide variety of mentalising tasks – involving the spontaneous attribution of mental states to animated shapes, reflecting on one's intentions to carry out certain actions, thinking about the preferences and dispositions of oneself or a fictitious story character, and judging the sincerity or sarcasm of another person's communicative intentions. Despite the variety of mentalising tasks used, these studies of mental state attribution have consistently shown that mPFC cortex activity during mentalising tasks decreases between adolescence and adulthood (Figure 1). Each of these studies compared brain activity in young adolescents and adults while they were performing a task that involved thinking about mental states (see Figure 1 for details of studies). In each of these studies, mPFC activity was greater in the adolescent group than in the adult group during the mentalising task compared to the control task. In addition, there is evidence for differential functional connectivity between mPFC and other parts of the mentalising network across age.[9]

To summarize, a number of developmental neuroimaging studies of social cognition have been carried out by different labs around the world, and there is striking consistency with respect to the direction of change in mPFC activity. It is not yet understood why mPFC activity decreases between adolescence and adulthood during mentalising tasks, but two non-mutually exclusive explanations have been put forward (see Blakemore[10] for details). One possibility is

Downloaded from jrs.sagepub.com at GEORGETOWN UNIV MED CTR on May 25, 2015

Figure 1

A section of the dorsal MPFC that is activated in studies of mentalising: Montreal Neurological Institute (MNI) 'y' coordinates range from 30 to 60, and 'z' coordinates range from 0 to 40. Dots indicate voxels of decreasing activity during seven mentalising tasks between late childhood and adulthood.[10,16] The mentalising tasks ranged from understanding irony, which requires separating the literal from the intended meaning of a comment, thinking about one's own intentions, thinking about whether character traits describe oneself or another familiar other, watching animations in which characters appear to have intentions and emotions, and thinking about social emotions such as guilt and embarrassment[9]



Blakemore et al.[17] intentions understanding
Burnett et al.[9] social emotion
Wang et al.[18] irony comprehension
Moriguchi et al.[19] mentalising animations
Pfeifer et al.[20] self/other personality traits
Pfeifer et al.[21] self appraisals
Sebastian et al.[22] emotional ToM
Gunther Moor et al.[23] reading the mind in the eyes
Van den Bos et al.[24] in press: reciprocity

that the cognitive strategy for mentalising changes between adolescence and adulthood. A second possibility is that the functional change with age is due to neuroanatomical changes that occur during this period. Decreases in activity are frequently interpreted as being due to developmental reductions in grey matter volume, presumably related to synaptic pruning. However, there is currently no direct way to test the relationship between number of synapses, synaptic activity

Downloaded from jrs.sagepub.com at GEORGETOWN UNIV MED CTR on May 25, 2015

and neural activity as measured by fMRI in humans (see Blakemore[10] for discussion). If the neural substrates for social cognition change during adolescence, what are the consequences for social cognitive behaviour?

Online mentalising usage is still developing in mid-adolescence

Most developmental studies of social cognition focus on early childhood, possibly because children perform adequately in even quite complex mentalising tasks at around age four. This can be attributed to a lack of suitable paradigms: generally, in order to create a mentalising task that does not elicit ceiling performance in children aged five and older, the linguistic and executive demands of the task must be increased. This renders any age-associated improvement in performance difficult to attribute solely to improved mentalising ability. However, the protracted structural and functional development in adolescence and early adulthood of the brain regions involved in theory of mind might be expected to affect mental state understanding. In addition, evidence from social psychology studies shows substantial changes in social competence and social behaviour during adolescence, and this is hypothesized to rely on a more sophisticated manner of thinking about and relating to other people – including understanding their mental states.

Recently, we adapted a task that requires the online use of theory of mind information when making decisions in a communication game, and which produces large numbers of errors even in adults.[11] In our computerized version of the task, participants view a set of shelves containing objects, which they are instructed to move by a 'Director', who can see some but not all of the objects (Figure 2).[12] Correct interpretation of critical instructions requires participants to use the director's perspective and only move objects that the director can see (the director condition). We tested participants aged between 7 and 27 years and found that, while performance in the director and a control condition followed the same trajectory (improved accuracy) from mid-childhood until mid-adolescence, the mid-adolescent group made more errors than the adults in the director condition only. These results suggest that the ability to take another person's perspective to direct appropriate behaviour is still undergoing development at this relatively late stage.

Implications for society

There are many questions that remain to be investigated in this new and rapidly expanding field. The study of neural development during adolescence is likely to have important implications for society in relation to education and the legal treatment of teenagers, as well as a variety of mental illnesses that often have their onset in adolescence.[13]

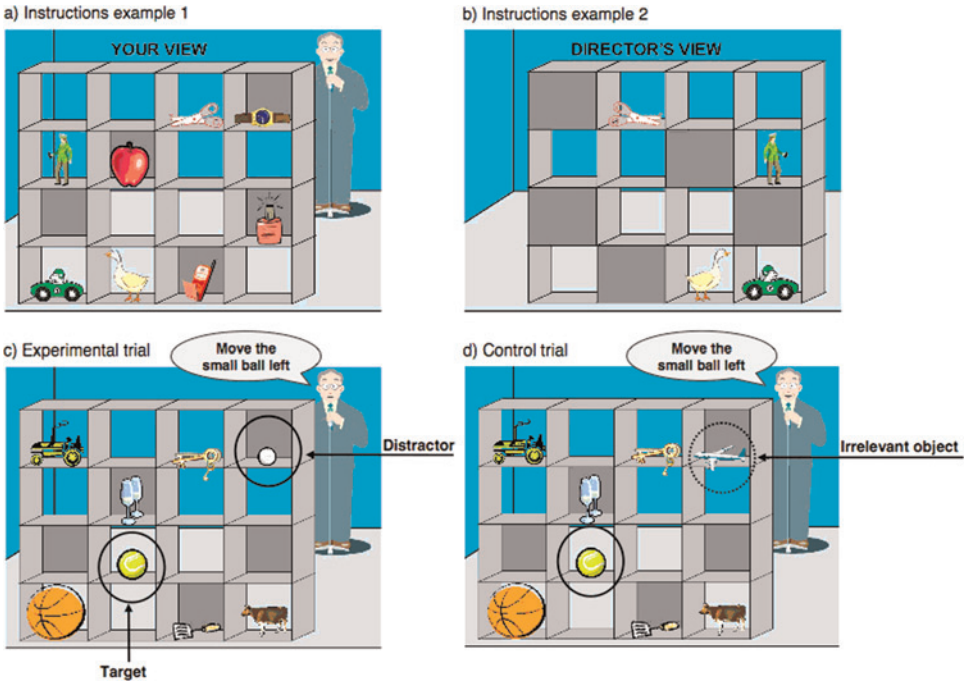
Research into the cognitive implications of continued brain maturation beyond childhood may be relevant to the social development and educational attainment of adolescents. Further studies are necessary to reach a consensus about how brain development impacts on social, emotional, linguistic, mathematical and creative development. In other words, which skills undergo perturbation, which undergo sensitive periods for enhancement and how does the quality of the environment interact with brain changes in the development of cognition? Longitudinal studies of the effect of early deprivation on the cognitive development of Romanian adoptees in the UK have begun to investigate this question.[14] Whether greater emphasis on social and emotional cognitive development would be beneficial during adolescence is unknown but research will provide insights into potential intervention schemes in secondary schools, for example, for remediation programs for antisocial behaviour.

Understanding the brain mechanisms that underlie learning and memory, and the effects of genetics, the environment, emotion and age on learning could transform educational strategies and enable us to design programs that optimize learning for people of all ages and of all needs. Neuroscience can now offer some understanding of how the brain learns new information and processes this information throughout life.[15] Understanding the brain basis of social functioning and social development is crucial to the fostering of social competence inside and outside the classroom. Social functioning plays a role in shaping learning and academic performance (and vice

Downloaded from jrs.sagepub.com at GEORGETOWN UNIV MED CTR on May 25, 2015

Figure 2

a, b. Images used to explain the Director condition: participants were shown an example of their view (a) and the corresponding director's view (b) for a typical stimulus with four objects in occluded slots that the director cannot see (e.g. the apple); c, d. Example of an experimental (c) and a control trial (d) in the Director condition. The participant hears the verbal instruction: 'Move the small ball left' from the director. In the experimental trial (c), if the participant ignored the director's perspective, she would choose to move the distractor ball (golf ball), which is the smallest ball in the shelves but which cannot be seen by the director, instead of the larger ball (tennis ball) shared by both the participant's and the instructor's perspective (target). In the control trial (d), an irrelevant object (plane) replaces the distractor item



versa), and understanding the neural basis of social behaviour can contribute to understanding the origins and process of schooling success and failure. The finding that changes in brain structure continue into adolescence (and beyond) has challenged accepted views, and has given rise to a recent spate of investigations into the way cognition (including social cognition) might change as a consequence. Research suggests that adolescence is a key time for the development of regions of the brain involved in social cognition and self-awareness. This is likely to be due to the interplay between a number of factors, including changes in the social environment and in puberty hormones, as well as structural and functional brain development and improvements in social cognition.

If early childhood is seen as a major opportunity – or a 'sensitive period' - for teaching, so too should the teenage years. During both periods, particularly dramatic brain reorganization is taking place. The idea that teenagers should still go to school and be educated is relatively new. And yet the research on brain development suggests that education during the teenage years is vital. The brain is still developing during this period, is adaptable, and needs to be moulded and shaped. Perhaps the aims of education for adolescents might change to include abilities that are controlled by the parts of the brain that undergo most change during adolescence. These abilities include internal control, multi-tasking and planning – but also self-awareness and social

Downloaded from jrs.sagepub.com at GEORGETOWN UNIV MED CTR on May 25, 2015

cognitive skills such as perspective taking and the understanding of other people's minds.

Conclusion

The study of brain development beyond childhood is a new but rapidly evolving field with applications for education and social policy. The finding that changes in brain structure and function continue into adolescence has challenged accepted views. In this paper, I have focused on research in developmental cognitive neuroscience, but a richer account of changes in adolescent learning, strategic and social behaviour requires a multidisciplinary approach that recognizes the complex interactions between genetics, brain structure, physiology and chemistry, and the environment.

References

1 Baillargeon R, Scott RM, Hea Z. False-belief understanding in infants. *Trends Cogn Sci* 2010;**14**:110–8
2 Barresi J, Moore C. Intentional relations and social understanding. *Behavioral and Brain Sciences* 1996;**19**:107–54
3 Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience* 2006;**7**:268–77
4 Burnett S, Blakemore SJ. Functional connectivity during a social emotion task in adolescents and in adults. *Eur J Neurosci* 2009;**29**:1294–301
5 Samson D, Apperly IA, Chiavarino C, Humphreys GW. Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience* 2004;**7**:499–500
6 Stuss DT, Gallup GG Jr, Alexander MP. The frontal lobes are necessary for 'theory of mind'. *Brain* 2001;**124**:279–86
7 Bird CM, Castelli F, Malik O, Frith U, Husain M. The impact of extensive medial frontal lobe damage on 'Theory of Mind' and cognition. *Brain* 2004;**127**:914–28
8 Steinberg L, Morris AS. Adolescent development. *Annual Review in Psychology* 2001;**52**:83–110
9 Burnett S, Bird G, Moll J, Frith C, Blakemore S-J. Development during adolescence of the neural processing of social emotion. *J Cogn Neurosci* 2009;**21**:1736–50
10 Blakemore SJ. The social brain in adolescence. *Nature Reviews Neuroscience* 2008;**9**:267–77
11 Keysar B, Lin S, Barr DJ. Limits on theory of mind use in adults. *Cognition* 2003;**89**:25–41
12 Dumontheil I, Apperly IA, Blakemore S-J. Online usage of theory of mind continues to develop in late adolescence. *Developmental Science* 2010;**13**:331–8
13 Paus T, Keshavan M, Giedd JN. Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience* 2008;**9**:947–57
14 Stevens SE, Sonuga-Barke EJ, Kreppner JM, *et al.* Inattention/overactivity following early severe institutional deprivation: presentation and associations in early adolescence. *J Abnormal Child Psychology* 2008;**36**:385–98
15 Blakemore SJ, Frith U. *The Learning Brain: Lessons for Education*. Oxford: Blackwell, 2005
16 Blakemore SJ. The developing social brain: implications for education. *Neuron* 2010;**65**:744–7
17 Blakemore SJ, den Ouden H, Choudhury S, Frith C. Adolescent development of the neural circuitry for thinking about intentions. *Soc Cogn Affect Neurosci* 2007;**2**:130–9
18 Wang AT, Lee SS, Sigman M, Dapretto M. Developmental changes in the neural basis of interpreting communicative intent. *Soc Cogn Affect Neurosci* 2006;**1**:107–21
19 Moriguchi Y, Ohnishi T, Mori T, Matsuda H, Komaki G. Changes of brain activity in the neural substrates for theory of mind during childhood and adolescence. *Psychiatry Clin Neurosci* 2007;**61**:355–63
20 Pfeifer JH, Lieberman MD, Dapretto M. "I know you are but what am I?!": neural bases of self- and social knowledge retrieval in children and adults. *J Cogn Neurosci* 2007;**19**:1323–37
21 Pfeifer JH, Masten CL, Borofsky LA, Dapretto M, Fuligni AJ, Lieberman MD. Neural correlates of direct and reflected self-appraisals in adolescents and adults: when social perspective-taking informs self-perception. *Child Dev* 2009;**80**:1016–38
22 Sebastian CL, Fontaine NMG, Bird G, *et al.* Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Soc Cogn Affect Neurosci* (in press)
23 Gunther Moor B, Op de Macks Z, Güroğlu B, Rombouts SARB, Van der Molen MW, Crone EA. Neurodevelopmental changes of reading the mind in the eyes. *Soc Cogn Affect Neurosci* (in press)
24 Van den Bos W, Van Dijk E, Westenberg PM, Rombouts SARB, Crone EA. Changing brains, changing perspectives: the neurocognitive development of reciprocity. *Psychol Sci* 2011;**22**:60–70

Downloaded from jrs.sagepub.com at GEORGETOWN UNIV MED CTR on May 25, 2015

Neuropsychopharmacology REVIEWS (2008) 33, 88–109
© 2008 Nature Publishing Group All rights reserved 0893-133X/08 $30.00

REVIEW

Stress, Depression, and Neuroplasticity: A Convergence of Mechanisms

Christopher Pittenger[1] and Ronald S Duman*,[1]

[1]*Department of Psychiatry, Connecticut Mental Health Center, Yale University School of Medicine, New Haven, CT, USA*

Increasing evidence demonstrates that neuroplasticity, a fundamental mechanism of neuronal adaptation, is disrupted in mood disorders and in animal models of stress. Here we provide an overview of the evidence that chronic stress, which can precipitate or exacerbate depression, disrupts neuroplasticity, while antidepressant treatment produces opposing effects and can enhance neuroplasticity. We discuss neuroplasticity at different levels: structural plasticity (such as plastic changes in spine and dendrite morphology as well as adult neurogenesis), functional synaptic plasticity, and the molecular and cellular mechanisms accompanying such changes. Together, these studies elucidate mechanisms that may contribute to the pathophysiology of depression. Greater appreciation of the convergence of mechanisms between stress, depression, and neuroplasticity is likely to lead to the identification of novel targets for more efficacious treatments.
Neuropsychopharmacology Reviews (2008) **33,** 88–109; doi:10.1038/sj.npp.1301574; published online 12 September 2007

Keywords: synapse; neurogenesis; antidepressant; signal transduction; gene expression

INTRODUCTION

Over the past 50 years, different categories of effective antidepressant medications have been fortuitously discovered. All available antidepressants act on the brain's modulatory monoamine systems, an observation that formed the core of the monoamine hypothesis of the pathophysiology of depression (Heninger *et al*, 1996). More recently, molecular events downstream of antidepressants' direct actions on the monoamines have been elucidated, generating new theories about the pathophysiology of depression and the action of antidepressant medications (Duman *et al*, 1997), and new families of potential targets for novel antidepressant therapies (Manji *et al*, 2003).

A startling observation, as these downstream molecular events have been elucidated, is the striking degree of overlap between the molecular and cellular changes induced by antidepressant treatment and the molecular mechanisms of neuroplasticity, especially synaptic plasticity (see review by Citri and Malenka in this volume). Perhaps the most well-characterized instance of this fact is the transcription factor CREB, which has a well-established role in learning-related synaptic plasticity in many organisms and brain regions (eg Bartsch *et al*, 1998; Pittenger *et al*, 2002; reviewed in Carlezon *et al*, 2005) and is involved in the hippocampus in antidepressant response (Thome *et al*, 2000; Chen *et al*, 2001a). Similar parallels have now been found in a multitude of other molecular events, synaptic alterations, and morphological changes, as we further review below. While the precise nature of the relationship between the pathophysiology of major depression and possible dysfunction of neuroplasticity remains poorly understood, it is likely that an intimate relationship exists.

In parallel, the interactions between chronic stress or a dysregulated stress response and the molecular, cellular, and behavioral changes that attend the development of a depression-like state have become increasingly clear. The relationship between psychosocial stressors and the development of depression in susceptible individuals has long been apparent (Kendler *et al*, 1999; Caspi *et al*, 2003). In experimental animals, stress can lead to atrophy of the hippocampus similar to that seen in depression (Sapolsky, 2000); chronic stress paradigms in animals recapitulate many of the core behavioral characteristics of depression and are responsive to antidepressant treatment (reviewed in Willner, 2005). The details remain to be elucidated, but any comprehensive picture of the pathophysiology of depression must include the role of stress in the etiology of the disorder.

Completing the circle, there is increasing appreciation of the effects of stress on the mechanisms of neuroplasticity (Shors *et al*, 1989; reviewed in McEwen, 1999). Clearly, then, there is an intimate relationship between the effects of stress, the mechanisms of neuroplasticity, and the pathophysiology of depression and mechanisms of antidepressant action.

*Correspondence: Dr RS Duman, Connecticut Mental Health Center, Yale University School of Medicine, 34 Park Street, Third Floor, New Haven, CT 6508, USA, Tel: +1 203 974 7726, Fax: +1 203 974 7724, E-mail: ronald.duman@yale.edu
Received 11 July 2007; revised 3 August 2007; accepted 4 August 2007

We explore this relationship in this review, outlining the major molecular and cellular pathways related to neuroplasticity that are altered by stress and appear to contribute to behaviors related to depression. We then discuss the mechanisms of antidepressant response, their overlap with mechanisms of neuroplasticity, and how they oppose the effects of chronic stress in various behavioral models. Together, these studies support a model of disruption of neuroplasticity by stress that contributes importantly to the pathophysiology of depression, and that is blocked or reversed by antidepressant treatment.

IMPAIRMENTS OF LEARNING AND MEMORY IN MAJOR DEPRESSION

Cognitive impairment is a core endophenotype of major depression (Hasler *et al*, 2004). One of the formal diagnostic criteria for the syndrome is a 'diminished ability to think or concentrate' (American Psychiatric Association, 2000), and patients often complain of difficulty with cognitive function during everyday tasks.

Cognitive difficulties in major depression fall into at least two domains, which are likely to correspond to different underlying disruptions of brain function. First, as indicated by the diagnostic criteria, is impairment of concentration and attention. Because the dorsolateral prefrontal cortex (DLPFC) is well established to play a critical role in these capacities, both in healthy humans and in experimental animals (Goldman-Rakic, 1996), this symptom is likely to relate to the well-documented abnormalities of DLPFC function in subjects with major depression (Baxter *et al*, 1989; Harvey *et al*, 2005) and to the finding of neuropathological change in this region in post-mortem tissue (Rajkowska *et al*, 1999, 2007).

Patients with major depression—both first episode and recurrent—also exhibit prominent deficits in explicit memory (Zakzanis *et al*, 1998), a cognitive capacity well established to depend on the function of the hippocampus and the medial temporal lobe (Squire *et al*, 2004). Hippocampal atrophy has been repeatedly documented in major depression: while the total number of neurons and glia has not been found to be altered, neurons are reduced in size and the volume of the neuropil is reduced (Stockmeier *et al*, 2004). Correspondingly, structural imaging has demonstrated decreased hippocampal size in patients with major depression, especially those who have suffered multiple episodes (MacQueen *et al*, 2003). It remains to be established whether reduced hippocampal size is the cumulative result of multiple major depressive episodes or whether it precedes the multiple episodes and represents a trait marker of vulnerability for recurrent disease; the two possibilities need not be mutually exclusive.

Disruption of hippocampal function, including the capacity for neuroplasticity, could contribute to several aspects of major depression. In addition to its clear role in declarative memory, the hippocampus is a key regulator of prefrontal cortical function; hippocampus and DLPFC function cooperatively to regulate explicit memory. Disruption of hippocampal function in major depression could therefore contribute to the observed deficits in concentration, described above. Hippocampal afferents are also critical regulators of both the nucleus accumbens and the ventral tegmental area (VTA). It has been hypothesized that an indirect excitatory projection from hippocampus to VTA is critical for coordinating the firing of VTA cells in response to novelty (Lisman and Grace, 2005); impairment of this hippocampal function could thus lead to reduced dopaminergic tone and contribute to anhedonia (Warner-Schmidt and Duman, 2006). Finally, the hippocampus provides an important source of negative modulation of the hypothalamus–pituitary–adrenal stress hormone axis through its projections to the hypothalamus; hippocampal dysfunction therefore may contribute to the dysregulation of the stress response that is seen in major depression.

STRESS EFFECTS ON NEUROPLASTICITY: A PATHOPHYSIOLOGICAL CONTRIBUTOR TO MAJOR DEPRESSION?

Chronic stress has many effects on the central nervous system, including effects on neuroplasticity in brain structures that are functionally abnormal in major depression. Given the clear clinical relationship between stress and major depression, these effects represent candidate pathophysiological links between stress, the mechanisms of neuroplasticity, and the development of major depressive disorder (MDD).

Hippocampus: Effects of Prolonged Stress on Hippocampus-Dependent Memory, Plasticity, Cell Survival, and Neurogenesis

Transient mild stress can enhance learning and memory (Luine *et al*, 1996). However, chronic or severe stress is decidedly disruptive of hippocampus-dependent memory in experimental animals (Conrad *et al*, 1996; de Quervain *et al*, 1998; Diamond *et al*, 1999; reviewed in Sapolsky, 2003). Extended or high-dose treatment with glucocorticoids has a similar effect (Bodnoff *et al*, 1995; de Quervain *et al*, 1998). Specific impairments of hippocampus-dependent explicit memory are also seen after treatment of human subjects with glucocorticoids (Newcomer *et al*, 1999; de Quervain *et al*, 2000) and after stress (reviewed in Shors, 2005).

Hippocampal synaptic plasticity, as modeled by long-term potentiation (LTP), is widely believed to represent an important component mechanism of hippocampus-dependent memory formation (Malenka and Bear, 2004; see Figure 1). It is therefore striking that a sufficiently severe stress can impair LTP in the rodent hippocampus (Foy *et al*, 1987; Shors *et al*, 1989; reviewed in Kim and Diamond, 2002). Conversely, similar stress paradigms in rodents enhance long-term depression (LTD) in the hippocampus (Xu *et al*, 1997); the two effects are likely to have overlapping mechanisms, as both are prevented by *N*-methy-D-aspartate (NMDA) blockade during the behavioral stress (Kim *et al*, 1996). The effect of serum glucocorticoids mirrors that of stress: low levels of glucocorticoids amplify LTP (perhaps through preferential activation of hippocampal mineralocorticoid receptors), while higher levels attenuate it (perhaps because they saturate the mineralocorticoid receptors and lead to robust activation of glucocorticoid receptors; reviewed in De Kloet, 2004).

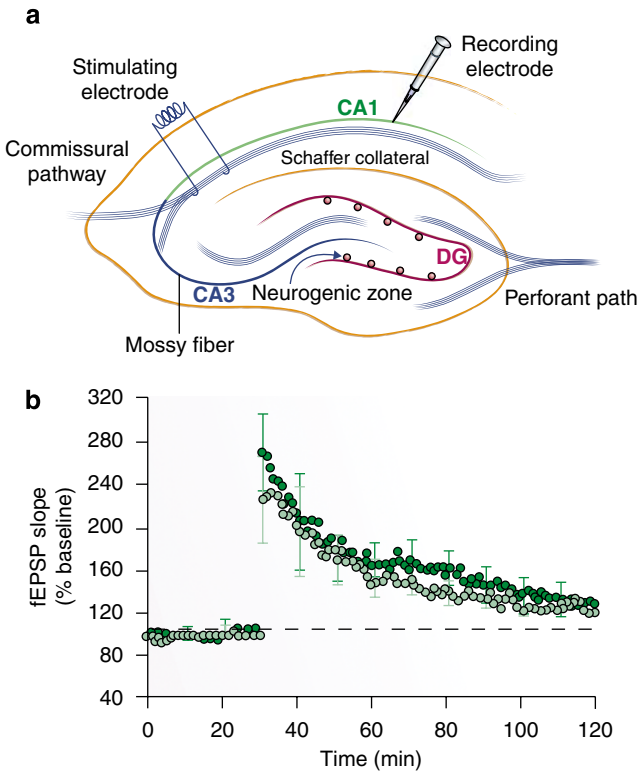


Figure 1 Hippocampal anatomy and sites of well-characterized forms of neuroplasticity. Neuroplasticity has been particularly intensively studied in the hippocampus. (a) Anatomy of the rodent hippocampus in coronal section is shown. Major synaptic projections include the perforant path from entorhinal cortex to dentate gyrus, the mossy fiber pathway from dentate gyrus to area CA3, and the Schaffer collateral pathway from area CA3 to area CA1. Other pathways exist but have been less intensively studied and are left out for clarity. Neurogenesis occurs in the subgranular cell zone of the dentate gyrus. (b) Synaptic plasticity has been characterized in all major synaptic pathways, but has been particularly intensively studied in the Schaffer collateral pathway; typical placement of stimulating and recording electrodes in an *in vitro* LTP experiment is shown.

Sustained levels of stress or glucocorticoids also damage the hippocampus at the level of morphological neuroplasticity (reviewed in Sapolsky, 2000) (Figure 2). Excess glucocorticoids (Woolley *et al*, 1990) or behavioral stress (Watanabe *et al*, 1992; Magarinos *et al*, 1996) lead to atrophy and retraction of the apical dendrites of hippocampal pyramidal cells; this effect leads to a reduction in the amount of neuropil without frank cell loss, reminiscent of what has been documented in the post-mortem hippocampus of patients with major depression (Stockmeier *et al*, 2004). Prolonged high-dose corticosterone—at higher levels than are typically achieved *in vivo*—can even result in death of hippocampal pyramidal cells (Sapolsky *et al*, 1985).

A final mechanism whereby prolonged stress can negatively impact hippocampal function and capacity for neuroplasticity has come to light more recently, with the broad acceptance of the presence of neurogenesis in the adult hippocampus (Figure 2). Many different forms of acute and chronic stress have been shown to reduce neurogenesis in the rodent hippocampus (reviewed in Duman, 2004; Dranovsky and Hen, 2006). Elevated levels of glucocorticoids likewise suppress hippocampal neurogenesis (Gould *et al*, 1992). As neurogenesis appears to be required for the behavioral response to antidepressants in rodents (Santarelli *et al*, 2003; see below) and impaired neurogenesis has been hypothesized to represent a core pathophysiological feature of major depression (Duman, 2004), this effect represents yet another way that the effects of stress on mechanisms of neuroplasticity may contribute to the development of depression.

Prefrontal Cortex: Effects of Prolonged Stress on Morphology and Function

While synaptic and morphological plasticity have been less intensively studied in the prefrontal cortex (PFC) than in the hippocampus, it is increasingly clear that stress has

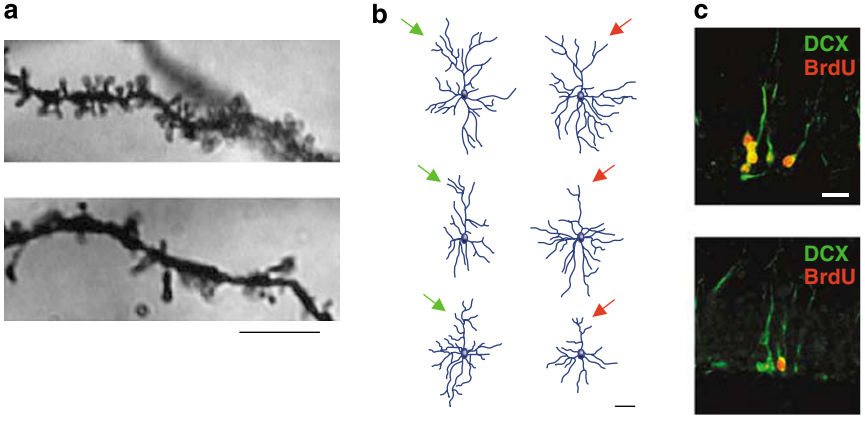


Figure 2 Stress alters neuroplasticity at multiple structural levels. (a) Chronic stress can reduce the number of dendritic spines. In this example, 8 weeks of postweaning social isolation produced a significant reduction in dendritic spines in mPFC in stress rat (lower panel) compared to controls (upper panel), as visualized by Golgi-Cox staining. From Silva-Gomez *et al*, 2003, with permission; scale bar = 5 μm. (b) Chronic stress can reduce the length and complexity of coritcal dendrites. In this example, repeated restraint stress (3 h a day for 3 weeks) produced shorter and less complex apical dendrites (indicated by arrows) in stressed rats (on the right) as compared to controls (on the left), as visualized in mPFC pyramidal neurons by computer-assisted resconstructions from Golgi-stained material. From Cook and Wellman, 2004, with permission; scale bar = 50 μm. (c) Chronic stress can impair neurogenesis. In this example, 3 weeks of daily unpredictable stressors reduced the generation of new neurons in the rat dentate gyrus (lower panel) compared to unstressed controls (upper panel). Neurogenesis is visualized by labeling with bromo-deoxyuridine (BrdU), which labels recently divided cells, and doublecortin (DCX), a marker of immature neurons. Ja-Wook Kim and RSD; scale bar = 20 μm.

similar effects on the mechanisms of neuroplasticity there. Chronic restraint stress induces significant regression of the apical dendrites of pyramidal cells in medial prefrontal cortex (mPFC) in rats, an effect similar to that described in area CA3 of the hippocampus (Cook and Wellman, 2004; Radley *et al*, 2004). The effect may be specific to mPFC, as orbital PFC is spared (Liston *et al*, 2006). Attentional set-shifting, a behavioral task that depends on intact mPFC function, is impaired in chronically stressed animals (Liston *et al*, 2006). A similar effect is seen following chronic administration of corticosterone (Wellman, 2001). Again, this morphological change appears to recapitulate some of the changes seen in post-mortem tissue from patients with MDD (Rajkowska *et al*, 1999).

One of the most consistent neuropathological findings in MDD is a reduction in the number of glia (Ongur *et al*, 1998; Rajkowska *et al*, 1999; Cotter *et al*, 2001, 2002; Uranova *et al*, 2004). In animals, chronic unpredictable stress results in a reduction in the proliferation of glia and endothelial cells in the mPFC (Banasr *et al*, 2007); exposure to glucocorticoids causes a similar effect (Alonso, 2000). Glia provide metabolic support for neurons; a reduction in the numbers of these cells could impact the function as well as morphology of mPFC pyramidal cells. Glia also play an important role in both the synthesis and inactivation of glutamate, which is central to many forms of neuroplasticity (a point to which we return below). Altered number or function of glia could thereby impact neuroplasticity. Stress-induced reductions in glial proliferation could contribute to the decrease in glial number observed in MDD and to a decrease in neural plasticity.

It is tempting to speculate that such stress-induced atrophy of prefrontal dendrites and a reduction in glial number contribute to the 'hypofrontality' observed in patients with major depression (see above). Some data support an effect of chronic stress on prefrontal physiology and information processing. For example, acute stress can perturb synaptic plasticity at the projection from amygdala to PFC (Maroun and Richter-Levin, 2003). At the reverse projection, from PFC to amygdala, stress shifts the balance from one that favors LTD to one that favors LTP (Maroun, 2006).

While the precise contribution of such perturbations to the network changes in brain function that characterize the depressed state is difficult to infer, these findings support the notion that sufficient levels of stress alter the mechanisms of neuroplasticity in a group of interconnected structures, which are functionally abnormal in major depression. Furthermore, even though only a few studies have examined synaptic plasticity in these structures after stress, the directions of the documented changes are concordant with the abnormalities seen in major depression. PFC is hypoactive in major depression; and potentiation of a major excitatory pathway leading to it is attenuated by stress. Conversely, the amygdala is hypertophic and hyperactive in major depression, and potentiation of a major input to it—from the ventro mPFC —is enhanced.

Amygdala: Stress-Induced Hypertrophy

Whereas the hippocampus and PFC are both reduced in size and activity in major depression, the amygdala's size and activity are increased (reviewed in Drevets, 2003). Several structural imaging studies have reported increased amygdala volume in patients with major depression (Bremner *et al*, 2000; Frodl *et al*, 2002; Lange and Irle, 2004). The amygdala has also been found to be hyperactive in major depression (Drevets *et al*, 1992); its activity correlates with the intensity of negative affect (Drevets *et al*, 1992; Abercrombie *et al*, 1998).

Chronic stress can enhance amygdala-dependent learning, in contrast to its effects on hippocampus-dependent declarative learning. In rats, chronic stress enhances amygdala-dependent fear learning (Conrad *et al*, 1999). It also enhances behavioral measures of anxiety in experimental animals (Conrad *et al*, 1999; Vyas *et al*, 2004).

Correspondingly, stress enhances synaptic plasticity and the function of amygdala neurons, an effect quite distinct from the atrophy it induces in the hippocampus and PFC. This could both result from and contribute to over-activation of neuronal circuits that control fear, anxiety, and emotion. In rats, stress that leads to atrophic changes in hippocampal pyramidal cells produces enhanced dendritic length and branching in amygdala principal cells and in the bed nucleus of the stria terminalis, an important amygdala projection target (Vyas *et al*, 2002, 2003). Prolonged stress increases dendritic spines and synaptic connectivity in the amygdala (Vyas *et al*, 2006). Interestingly, these changes do not reverse even several weeks after the cessation of the chronic stressor (Vyas *et al*, 2004). Indeed, the effects of stress on the amygdala maybe very long-lasting indeed: in rats, prenatal stress can lead to increased amygdala size in adulthood (Salm *et al*, 2004). If replicated and extended, such findings have clear implications for the pathophysiological relationship between early life stress and subsequent vulnerability to major depression.

The contrast between the effects of stress on the amygdala and those on the hippocampus and PFC is striking. Whereas acute and chronic stress impair hippocampal function, reduce the length and complexity of CA3 dendrites, and impair neurogenesis, very similar treatments lead to enhanced amygdala-dependent fear learning, increased length and complexity of amygdalar dendrites, and increased amygdala size. This contrast makes it clear that the well-documented effects of stress on hippocampal morphology and function are not manifestations of a universal effect of stress hormones or other aspects of stress on neuronal integrity. Rather, the effects of stress on brain morphology and function are region and circuit dependent. Both acute and chronic stress can have very different effects of different brain regions and functions, a fact with profound implications for our understanding of the pathophysiology of major depression.

Ventral Striatum: Stress-Associated Changes in Neuroplasticity and the Mechanisms of Reward

A final structure in which neuroplasticity may relate to effects of stress and the symptoms of depression is the ventral striatum, including the nucleus accumbens. The accumbens has a central role in the mechanisms of natural reward; as such, its dysregulation in depression is thought to relate to symptoms of anhedonia (Dunn *et al*, 2002; Nestler and Carlezon, 2006). A recent neuroimaging study

supports this view, finding attenuated accumbens activation in response to positive visual stimuli in depressed subjects (Epstein *et al*, 2006). The relevance of such dysfunction, both to pathophysiology and to future therapies, was dramatically demonstrated by the successful treatment of several cases of profoundly refractory depression by deep brain stimulation of the nucleus accumbens (Schlaepfer *et al*, 2007).

Stress can activate the dopaminergic projection to the accumbens from the VTA; this may contribute to a homeostatic response to stress or to adaptive stress-related learning. However, chronic stress can cause long-term adaptations in the VTA–accumbens pathway (Ortiz *et al*, 1996; Saal *et al*, 2003) that may contribute to its dysregulation in major depression (reviewed in Nestler and Carlezon, 2006). Though plasticity in this pathway under conditions of stress and depression remains sparsely studied, it is likely to represent an important component of any comprehensive view of the relationship between stress, depression, and neuroplasticity.

MOLECULAR AND CELLULAR MECHANISMS OF NEUROPLASTICITY

The mechanisms of synaptic and morphological plasticity have been extensively studied in the context of their contribution to learning and memory (Malenka and Bear, 2004; Citri and Malenka, this volume). It is here that some of the most striking parallels between the mechanisms of neuroplasticity and those underlying MDD emerge, especially in the overlap between the cellular mechanisms of synaptic plasticity and the molecular and cellular changes that attend antidepressant treatment. A dizzying array of molecular components has been implicated in synaptic plasticity in various systems (Sanes and Lichtman, 1999); it is not our intention to comprehensively review this vast literature. Rather, we emphasize some of the common themes in studies of synaptic and structural plasticity, and some of the specific molecular players most cogent to a subsequent discussion of the mechanisms of antidepressant response.

Inducers and Local Mechanisms of Synaptic Plasticity

Mechanistically distinct forms of experimentally induced synaptic plasticity have been described in a number of different systems (Malenka and Bear, 2004). A common theme is that many forms of synaptic potentiation are triggered by increases in synaptic calcium influx and in the local concentration of the second messenger molecule cyclic AMP (cAMP). These pathways are particularly well suited to the requirements of a plasticity-inducing mechanism. Local calcium influx, for example, commonly derives from the NMDA receptor, which is a highly evolved coincidence detector—it is activated only when presynaptic and postsynaptic cells are depolarized simultaneously. Such coincidence detection is precisely what is required for many forms of homosynaptic plasticity (Hebb, 1949). cAMP is regulated by many modulatory neurotransmitters, including serotonin, dopamine, and norepinephrine, as well as by calcium; it is therefore optimally suited for the integration of synaptic events with the modulatory influences of global variables such as arousal and attention.

Local elevations in calcium and cAMP induce events required for short-term synaptic plasticity. For example, the calcium-calmodulin-dependent kinase II (CaMKII), which is activated by local increases in calcium, is critical for early LTP. CaMKII has unique properties that make it well suited to a role in the induction of synaptic plasticity. Upon induction by a sufficient rise in the local concentration of calcium, it can phosphorylate itself. Autophosphorylated CaMKII is persistently active, even after calcium levels fall. This mechanism allows an inducing pulse of calcium to lead to more persistent activation of CaMKII, and thereby to activation of downstream processes that result in synaptic change (Lisman and Goldring, 1988).

Several other calcium-calmodulin kinases exist in neurons. CaMKIV, in particular, is prominent in the neuronal nucleus, where it is an important activator of regulated transcription factor such as CREB (Bito *et al*, 1996). Inhibition of CaMKIV impairs both long-lasting forms of LTP and long-lasting hippocampus-dependent learning (Kang *et al*, 2001). Calcium calmodulin-dependent kinases can thus contribute to both short-term and long-lasting neuroplasticity.

Both CaMKII and other kinases can phosphorylate the GluR1 subunit of the AMPA glutamate receptor and associated proteins. This phosphorylation both increases the number of AMPA receptors in the postsynaptic membrane (by triggering insertion of new AMPA receptors) and enhances the function of those receptors already inserted. Both mechanisms contribute to an enhancement of synaptic strength.

At the structural level, insertion of AMPA receptors can result in the activation of 'silent synapses' (Malenka and Nicoll, 1997). A silent synapse is a synaptic connection at which the postsynaptic membrane contains only NMDA receptors (Figure 3). Because of the NMDA receptor's coincidence detection properties, such a synapse is not activated by simple depolarization of the presynaptic cell. Upon the induction of LTP, AMPA receptors are inserted into the postsynaptic membrane at such a synapse, rendering it active upon subsequent single presynaptic impulses (hence no longer 'silent'). This unsilencing of silent synapses contributes to synaptic strengthening and appears to be a major mechanism of LTP.

Transduction to the Nucleus

While short-term changes in synaptic strength can be mediated by local events at the synapse, such as AMPA receptor phosphorylation and insertion into the postsynaptic membrane, longer-term changes require a broader coordination of cellular mechanisms—in particular, induction of genes and production of new protein (Nguyen *et al*, 1994). Upon sufficient local elevation of calcium and cAMP, therefore, signal transduction cascades are activated that transmit the inducing signal to the nucleus (Kandel, 2001). (Specific signaling molecules, such as kinases and phosphatases, can participate both in local, early events and in signaling to the nucleus and the transition to more

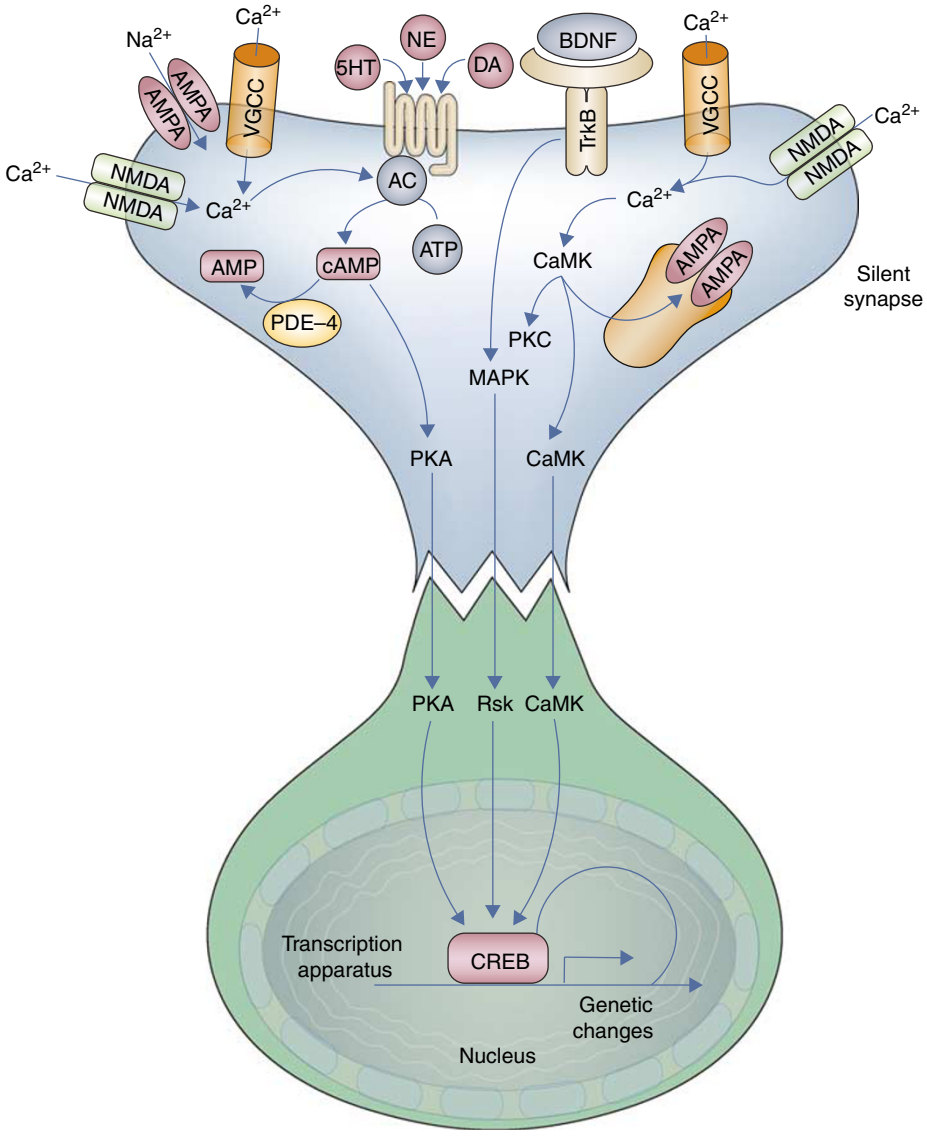


Figure 3 Major molecular pathways involved in neuroplasticity and affected by stress, depression, and antidepressant treatment. Some major molecular pathways involved in both short- and long-term neuroplastic changes are shown. Certain intermediates and other details are left out for clarity. Many of these pathways are influenced in opposite ways by stress and depression. For example, both chronic stress in animals and depression in humans have been associated with reductions in the transcription factor CREB, and antidepressants enhance CREB activity in the hippocampus. See text for further details and references. Abbreviations: NMDA, *N*-methyl-D-aspartate glutamate receptor; AMPA, amino-3-hydroxy-5-methyl-isoxazole-4-propionic acid glutamate receptor; VGCC, voltage-gated calcium channel; 5-HT, 5-hydroxytryptamine (serotonin); NE, norepinephrine; DA, dopamine; BDNF, brain-derived neurotrophic factor; Trk-B, BDNF receptor; AC, adenylyl cyclase; ATP, adenosine triphosphate; cAMP, cyclic adenosine monophosphate; AMP, adenosine monophosphate; PDE, phosphodiesterase; CaMK, calcium-calmodulin-dependent kinase; PKC, protein kinase C; MAPK, mitogen-activated protein kinase; PKA, protein kinase A; Rsk, ribosomal S6 protein kinase; CREB, cAMP response element-binding protein.

persistent forms of plasticity, as in the case of CaM kinases discussed above).

One signal transduction cascade that has been particularly clearly tied to the mechanisms of late-phase LTP (L-LTP) is the cAMP-dependent protein kinase, PKA (Abel *et al*, 1997). PKA consists of a regulatory and a catalytic subunit; at rest, the regulatory subunit binds to the catalytic subunit, inactivating it. Upon sufficient accumulation of cAMP, the two subunits dissociate, freeing the catalytic subunit to phosphorylate a variety of substrates. Since cAMP is increased by β-adrenergic, 5-HT4, 6, 7, and dopamine D1 receptors, activation of PKA represents a mechanism by which modulatory neurotransmitters can influence the mechanisms of long-lasting synaptic plasticity.

Another signal transduction pathway that has been repeatedly implicated in signaling to the nucleus and the induction of L-LTP is the mitogen-activated protein kinase (MAPK) pathway. This cascade of kinases is critical for the induction of long-lasting synaptic plasticity in the hippocampus, amygdala, and cortex (Atkins *et al*, 1998; Huang *et al*, 2000; Di Cristo *et al*, 2001; reviewed in Giovannini, 2006). MAPK translocates to the nucleus upon induction of L-LTP, where it can activate nuclear substrates (Patterson *et al*, 2001).

The induction of new genes in support of L-LTP requires that such signaling systems activate regulated transcription factors in the nucleus. There are likely to be many such regulated transcription factors involved in coordinating the genes that contribute to long-lasting synaptic change. Particularly, strong evidence implicates the transcription factor CREB in the regulation of numerous forms of synaptic change: it is activated upon synaptic stimulation and after learning in the hippocampus, amygdala, and cortex; and CREB inhibition in these and other brain regions disrupts LTP and corresponding forms of long-lasting memory (Glazewski *et al*, 1999; Pittenger *et al*, 2002, 2006; Kida *et al*, 2002; reviewed in Carlezon *et al*, 2005). CREB can be activated by a variety of kinases, including PKA and (indirectly) the MAPK cascade.

Effector Genes—Neurotrophins and Structural Change

CREB, and other inducible transcription factors, induce effector genes that contribute to the stabilization of synaptic plasticity (Kandel, 2001). Prominent among these is brain-derived neurotrophic factor (BDNF), which is induced by LTP and has a critical role in stabilizing synaptic change (Patterson *et al*, 1992). Once again, this has been demonstrated in hippocampus (Patterson *et al*, 1996), amygdala (Rattiner *et al*, 2005), and cortex (Bartoletti *et al*, 2002). Knockout of BDNF is lethal because of its multiple developmental roles, complicating analysis of its role in learning and memory; but recent studies that have disrupted BDNF only in the adult animal indicate a critical role in information processing and storage (Monteggia *et al*, 2004; Heldt *et al*, 2007; reviewed in Pang and Lu, 2004).

BDNF acts by multiple mechanisms and influences both early and late phases of synaptic plasticity, in both the presynaptic and the postsynaptic cells. It acts, at least in part, via the MAPK signaling cascade, suggesting that this pathway plays multiple roles in the regulation of plasticity. Other growth factors have also been demonstrated to influence LTP, including vascular endothelial growth factor (VEGF; Cao *et al*, 2004). Like BDNF, VEGF activates the MAPK cascade and has been implicated in the actions of stress and antidepressant treatments, as will be further discussed below.

It has long been hypothesized that long-lasting synaptic change is likely to correspond to morphological change at potentiated (or depressed) synapses. In recent years, growth of new dendritic spines and enlargement of existing spines have been demonstrated after LTP-inducing synaptic stimulation (Engert and Bonhoeffer, 1999; Matsuzaki *et al*, 2004). As a neurotrophic factor, with a critical role in stabilizing neurons during development, BDNF is well equipped to participate in such changes. The same is true for other LTP-induced secreted factors such as the tissue plasminogen activator (Pang and Lu, 2004) and cell adhesion molecules such as NCAM (Bukalo *et al*, 2004).

Positive and Negative Regulators of Neuroplasticity

As is expected of any important physiological process, synaptic potentiation and other forms of neuroplasticity are controlled by both positive and negative regulatory mechanisms (Abel *et al*, 1998). LTD may under some circumstances provide a homeostatic counterbalance to excessive synaptic potentiation (Tononi and Cirelli, 2006). Inhibitors of the signal transduction cascades that contribute to synaptic potentiation also provide a counterbalancing influence. Examples include phosphatases such as calcineurin (Malleret *et al*, 2001), which can antagonize signaling through the MAPK cascade, and phosphodiesterases, which break down cAMP and thus attenuate PKA-mediated signaling and other cAMP-dependent processes (Barad *et al*, 1998).

Recent studies have revealed that the NMDA receptor, which is the canonical initiator and positive regulator of LTP, can have contrary effects depending on its subcellular localization. Specifically, while synaptically localized NMDA receptors can trigger LTP, positively regulate MAPK, CREB and BDNF, and contribute to long-lasting plasticity and cell survival, activation of extrasynaptic NMDA receptors has the opposite effects, inhibiting CREB activation and the production of BDNF (Hardingham *et al*, 2002). It has been proposed that robust activation of extrasynaptic NMDA receptors may be an important contributor to excitotoxic cell death (Hardingham and Bading, 2003). Imbalance between stimulation of synaptic and extrasynaptic receptors may contribute to pathological states, including depression (as will be further explored below; reviewed in Pittenger *et al*, 2007).

Neurogenesis

Contrary to long-standing dogma, clear evidence now demonstrates that new neurons are generated in the adult mammalian brain; neurogenesis is prominent in the dentate gyrus region of the hippocampal formation (Altman and Das, 1965; Kaplan and Hinds, 1977; Kempermann *et al*, 1997; reviewed in Ming and Song, 2005).

The specific functional role and relevance of these new neurons is less firmly established; but a link between neurogenesis and the learning-related functions of the hippocampus is an intriguing possibility. Several lines of evidence support such a functional link (Leuner *et al*, 2006). Some computational theories of hippocampal function predict a role for new neurons in hippocampus-dependent learning (Chambers *et al*, 2004). Performance in a hippocampus-dependent spatial learning task correlates with the number of new neurons in aged rats (Drapeau *et al*, 2003), suggesting a functional contribution to learning. Electrical activity has been shown to be coupled with neurogenesis in the hippocampus, suggesting a mechanism whereby recruitment of the hippocampus during learning may lead to the production and incorporation of new neurons into the circuit (Deisseroth *et al*, 2004).

Behavioral studies in animals in which neurogenesis has been experimentally perturbed lend support to the idea that the new cells play a critical causal role in some forms of hippocampus-dependent learning. Administration of a chemotherapeutic agent that kills dividing cells impairs step-down avoidance, which depends on the function of the hippocampus (along with other structures; Shors *et al*, 2001). Saxe *et al* (2006) found that ablation of new neurons impaired dentate gyrus LTP (as measured *in vitro*) and

hippocampus-dependent contextual fear conditioning, but did not affect cued fear conditioning (which does not require the hippocampus). Data on the role of the new neurons in spatial learning are conflicting: Saxe *et al* (2006) found ablation of the new neurons to spare spatial learning in the Morris water maze, while others have found interference with neurogenesis to compromise spatial learning (Snyder *et al*, 2005). Further studies will be needed to clarify the role that new hippocampal neurons play in learning and memory.

Mechanisms initially implicated in the induction of LTP are likely also to be important in the regulation of neurogenesis. For example, activation of CREB promotes neurogenesis (Nakagawa *et al*, 2002a), and blockade of CREB function decreases neurogenesis (Nakagawa *et al*, 2002a, b). Likewise, growth factors regulate both the proliferation (VEGF; Cao *et al*, 2004) and survival (BDNF; Sairanen *et al*, 2005) of new hippocampal neurons.

This outline of molecular pathways contributing to neuroplasticity is necessarily selective and incomplete. More thorough treatments are given elsewhere (Malenka and Bear, 2004; Citri and Malenka, this volume), although this literature is so vast that it has become virtually impossible to review comprehensively. It does, however, lay the groundwork for an examination of how behavioral stress and pharmacological treatment with glucocorticoids alter signaling pathways associated with neuroplasticity.

IMPACT OF STRESS ON THE MOLECULAR PATHWAYS OF NEUROPLASTICITY

We have reviewed above some of the evidence that stress can lead to alterations in morphological plasticity: regression of dendrites, reduction in spine density, and a shrinkage of the neuropil in hippocampus and PFC. We now turn to effects of stress on some of the molecular mechanisms underlying neuroplasticity.

As noted above, acute and chronic stress can have quite different effects on learning: while acute stress can potentiate learning (and LTP; Shors, 2001), chronic stress leads to deficits in hippocampus-dependent memory reminiscent of those seen in major depression. It is therefore to be expected that the downstream molecular changes resulting from stress or manipulations of stress hormones may depend sensitively on the intensity and duration of the stressor glucocorticoid excess. This issue has not yet been explored in sufficient experimental detail for firm conclusions to be drawn. In the brief review presented here, we merely point out evidence that stress and glucocorticoids, either acute or chronic, can influence mechanisms implicated in neuroplasticity, without attempting to draw firm conclusions about the specific functional consequences of specific alterations in plasticity mechanisms or about mechanistic distinctions that may exist between the effects of acute and chronic stress on these systems.

Contributions of Glutamate to Neuronal Atrophy after Stress

As reviewed above, chronic stress or elevations in glucocorticoids can lead to neuronal atrophy, especially to dendritic retraction in cells of the CA3 cell field and the mPFC. Convergent evidence suggests that glutamatergic excess is likely to contribute to this cell damage and even, in extreme cases, to cell death (Sapolsky, 2000, 2003). Acute stressors rapidly increase extracellular glutamate in the PFC (Bagley and Moghaddam, 1997). Glucocorticoid excess increases glutamate release in the CA1 region of the hippocampus (Venero and Borrell, 1999), and chronic behavioral stress increases extracellular levels of glutamate in the CA3 region (Lowy *et al*, 1993). Glutamate antagonists can attenuate or block some of the effects of chronic glucocorticoid excess on dendritic morphology in the hippocampus (Magariños and McEwen, 1995). Exposure to glucocorticoids increases expression of the primary glial glutamate transporter, GLT-1 (Zschocke *et al*, 2005; Autry *et al*, 2006); this may serve to increase reuptake of elevated extracellular glutamate in conditions of chronic stress.

The finding of elevated extracellular glutamate in conditions of chronic stress is, at first blush, at odds with the suggestion that chronic stress attenuates mechanisms of neuroplasticity. This apparent contradiction is perhaps clarified by the recent distinction between the function of synaptic and extrasynaptic NMDA receptors (Hardingham *et al*, 2002; Hardingham and Bading, 2003). Stress-induced increases in extrasynaptic glutamate may perturb the balance between synaptic and extrasynaptic NMDA tone and have a net inhibitory effect on the mechanisms of synaptic plasticity and neuronal growth and survival (reviewed in Pittenger *et al*, 2007).

The effects of chronic stress on glutamatergic neurotransmission can be seen in alterations in the expression of glutamate receptor genes. For example, glucocorticoids can enhance AMPA-mediated excitatory synaptic transmission (Karst and Joëls, 2005). This contrasts with the finding that maternal separation stress has been shown to lead to lasting reduction in the expression of the NMDA receptor subunit NR2B and of the AMPA receptor subunits GluR1 and GluR2 (Pickering *et al*, 2006); differences in the level and chronicity of stress or glucocorticoid exposure may explain these contrasting results, which merit further study.

Stress-Induced Alterations Plasticity-Associated Signaling Pathways

Stress can alter neuronal signaling in multiple ways. Scattered observations in the literature suggest that chronic stress can alter signaling pathways implicated in synaptic plasticity. For example, several forms of chronic stress have been observed to increase the phosphorylation of MAPK (Pardon *et al*, 2005; Lee *et al*, 2006). Acute glucocorticoid treatment likewise induces the phosphorylation of MAPK (Revest *et al*, 2005), and acute swim stress leads to the phosphorylation of both MAPK and CaMKII (Ahmed *et al*, 2006). This activation of MAPK appears to be critical for the effects of behavioral stress on hippocampal LTP (Yang *et al*, 2004).

Both acute and chronic stressors can also lead to increased phosphorylation of CREB in the hippocampus (Pardon *et al*, 2005; Ahmed *et al*, 2006), consistent with an alteration in signaling pathways linking synaptic activity to nuclear effects. In contrast, chronic glucocorticoid treatment impairs CREB activity in cultured neurons (Focking

et al, 2003); certain chronic mild stress paradigms have been reported to produce a similar effect (Grønli *et al*, 2006). Again, these contrasting results may relate either to differences in experimental paradigm (Nair *et al*, 2007) or to a central role for the intensity or duration of stressors on the downstream molecular perturbations that result.

Stress Regulation of Growth Factors

As noted above, signaling pathways implicated in neuroplasticity target, among other downstream targets, genes for growth factors such as BDNF. Both acute and chronic stress lead to reductions in hippocampal BDNF mRNA levels, suggesting an impairment of some of the mechanisms of neuroplasticity (Nibuya *et al*, 1995, 1999; Smith *et al*, 1995; Russo-Neustadt *et al*, 2001; Rasmusson *et al*, 2002; Franklin and Perrot-Sinal, 2006). Glucocorticoids likewise suppress BDNF expression (Smith *et al*, 1995; reviewed in Schaaf *et al*, 2000).

Other growth factors are likewise regulated by stress. For example, nerve growth factor, the paradigmatic developmental neurotrophin, has been reported to be upregulated by chronic stress (Alfonso *et al*, 2006). VEGF, a trophic factor induced by electroconvulsive seizure (Newton *et al*, 2003), is suppressed by chronic stress (Heine *et al*, 2005). Similarly, VEGF's angiogenic actions are impaired by glucocorticoid treatment (Kasselman *et al*, 2007). The involvement of trophic factors beyond BDNF suggests that a multifaceted machinery of neuronal support may be impaired by stress and, possibly, enhanced by antidepressant therapies (Newton and Duman, 2004).

Other Downstream Target Genes

Recently, gene profiling of animals exposed to stress has identified other genes that are differentially regulated by stress or glucocorticoids (Alfonso *et al*, 2004); many genes so identified have roles in neuroplasticity. For example, one study found decreases in the cell adhesion molecule NCAM (whose upregulation is critical for long-lasting LTP; Muller *et al*, 1996), the signaling molecule PKC-α, and the synaptic marker synapsin I in various chronic stress models (Alfonso *et al*, 2006). Further characterization of neuroplasticity-related genes dysregulated by different stressors will provide opportunities to identify which particular neuroplasticity-related mechanisms are most closely associated with the effects of acute and chronic stress in the brain.

ANTIDEPRESSANTS INDUCE NEUROPLASTICITY

If the effects of stress on the mechanisms of neuroplasticity contribute to the pathophysiology of depression, then antidepressant treatments might be expected to affect the same mechanisms. Substantial evidence suggests that this is in fact the case. In the simplest scenario, where chronic stress impairs the mechanisms of neuroplasticity, antidepressants would be predicted to have the opposite effect and enhance them. There are multiple situations in which this is in fact the case. Other instances are likely to be more complicated, as stress and antidepressant treatment impact the complicated and interacting mechanisms of neuroplasticity in contrasting but not directly opposite ways. A deepened appreciation of these interactions will inform our evolving understanding of the pathophysiology of depression and the mechanisms of antidepressant action, and is likely to guide the search for novel antidepressant strategies in the future.

Enhanced Neuroplasticity and Cognitive Function with Treatment in Depressed Patients

If depression entails a deficit in neuroplasticity, then antidepressant treatments may enhance neuroplasticity and even reverse deficits produced during the symptomatic period. Few studies to date have examined the effect of antidepressant therapies on cognition in normal human subjects, in part simply because of the ethical difficulties that would attend administration of chronic antidepressant therapy to patients who are not ill. One recent double-blind study suggested improvement in memory and a variety of other cognitive domains with chronic fluoxetine treatment in elderly patients with mild cognitive impairment (Mowla *et al*, 2007). Given the mounting evidence that antidepressant treatment can promote neuroplasticity, further studies of this sort, using antidepressants with a primary goal of enhancing impaired cognitive function, may be warranted.

A small literature does address the improvement of cognitive deficits, and even of structural damage, in patients treated with a variety of antidepressant therapies. Electroconvulsive therapy (ECT) is perhaps the most efficacious antidepressant therapy (and, incidentally, the one with the most profound effects on the mechanisms of neuroplasticity). Examination of cognitive improvement in patients receiving ECT is complicated by the fact that memory difficulties, especially retrograde amnesia, are a clear side effect of the treatment (Sackeim, 2000). However, many patients show clear cognitive improvement, even shortly after treatment (Sackeim *et al*, 1992).

A few studies likewise show improvement in memory and other cognitive functions in depressed patients after pharmacological treatment. An early study found improvement in memory with treatment with the monoamine oxidase inhibitor (MAOI) moclobemide, which was superior to that of the adrenergic agonist viloxazine or the tetracyclic antidepressant maprotiline (Allain *et al*, 1992). The lack of anticholinergic effects of moclobemide was cited as a potentially important contributor to this difference; the anticholinergic effects of many older antidepressants, especially the tricyclics, may worsen cognitive function and mask any improvement produced by improvement of depression or enhancement of neuroplasticity. A more recent study found improvements in verbal memory after treatment of depressed patients with fluoxetine (Vythilingam *et al*, 2004); another found improvements in memory and attention in patients with late-onset depression after treatment with either fluoxetine or reboxetine (Gallassi *et al*, 2006). The absence of a placebo-treated control group makes it problematic to assign a causal role to the pharmacological treatment in these studies. More studies of the improvement in memory and other cognitive functions that may attend antidepressant treatment are clearly warranted.

As described above, stress can lead to atrophy of both hippocampus and PFC; morphological change in both regions is seen in patients with depression (MacQueen *et al*, 2003; Stockmeier *et al*, 2004). An ideal antidepressant treatment would both prevent this atrophy and, by stimulation of neurotrophic mechanisms, reverse it once it has occurred. No large longitudinal studies have been reported that might evaluate the ability of currently available antidepressants to accomplish this. Some studies examining hippocampal volume in patients with a long history of depression find a correlation with the amount of time spent symptomatically ill, suggesting that successful treatment and symptom reduction may halt a deteriorative process (Shah *et al*, 1998; Sheline *et al*, 1999; MacQueen *et al*, 2003). However, other studies have not found such a correlation (Bremner *et al*, 2000). There is one report that antidepressant treatment reverses the hippocampal volume reduction in PTSD patients and improves declarative memory (Vermetten *et al*, 2003). Additional longitudinal studies will be required to further examine the effects of antidepressant treatment on hippocampal volume in subjects with MDD.

Antidepressants Enhance Learning and Memory in Animal Models

Investigations of the effect of antidepressants on LTP have produced mixed results, with some studies reporting increases in LTP with antidepressant treatment, some decreases, and some with no effects. However, closer examination, comparing effects in subfields of hippocampus (ie dentate gyrus *vs* CA1 pyramidal cell layer) and/or types of antidepressants (tricyclic *vs* nontricyclic SSRIs), reveals more consistency.

In the dentate gyrus, where high-frequency stimulation of the perforant pathway projection to granule cells is used to stimulate LTP, both chronic ECS and chemical antidepressant treatment increase LTP (Stewart and Reid, 2000; Levkovitz *et al*, 2001). One of these studies (Stewart and Reid, 2000) found that chronic ECS or SSRI (fluoxetine) administration increases baseline field potentials in the dentate gyrus. *In vitro* LTP was reduced; this was interpreted as representing a ceiling effect, resulting from prior induction of plasticity by ECS or fluoxetine. The other study reports that either desipramine or mianserin administration results in increased LTP of dentate gyrus (Levkovitz *et al*, 2001). It is interesting to speculate that this increased potential for plasticity in the dentate could result from an increased number of newborn granule cells, which are known to have greater potential for neuroplasticity (Toni *et al*, 2007; Tashiro *et al*, 2007).

In contrast, most early studies of CA1 pyramidal cells report that chronic antidepressant treatment decreases LTP (Massicotte *et al*, 1993; O'Connor *et al*, 1993; Von Frijtag *et al*, 2001). However, these early studies were limited to tricyclic antidepressants (ie imipramine, triimipramine), which have substantial anticholinergic properties that could oppose other effects on plasticity. More recent reports, using antidepressant medications with less anticholinergic effect, have suggested that chronic administration of an SSRI or an atypical antidepressant (tianeptine) increases LTP and blocks the stress-induced impairment of LTP and enhancement of LTD in CA1 (Vouimba *et al*, 2006; Holderbach *et al*, 2007). Chronic SSRI administration has been reported to have similar effects on hippocampal-PFC circuits, reversing a stress-induced impairment of LTP and enhancement of LTD (Rocher *et al*, 2004). Together these studies indicate that chronic antidepressant treatment increases cellular plasticity in the dentate gyrus, and there is some evidence that SSRI antidepressants have a similar effect in on CA1, as well as block the effects of stress on this subpopulation of hippocampal neurons.

Much less is known about the influence of antidepressant treatments on behavioral models of learning and memory. Chronic administration of fluoxetine or venlafaxine, a mixed action reuptake inhibitor, is reported to improve performance in the Morris water maze, a spatial learning and memory model (Nowakowska *et al*, 2000, 2003, 2006), although another study found that fluoxetine did not influence performance in this task (Stewart and Reid, 2000). Chronic imipramine or tianeptine were not effective in the Morris water maze (Nowakowska *et al*, 2000, 2003), although another study found that chronic amitriptyline prevents age-induced impairments in learning and memory (Yau *et al*, 2002). This latter study also found a decrease in circulating corticosterone levels in antidepressant-treated rats that could contribute to the observed improvement in learning and memory. Tianeptine, but not fluoxetine, has been reported to improve discrimination performance in a radial maze task (Jaffard *et al*, 1991). Finally, chronic imipramine, but not paroxetine, has recently been reported to impair spatial working memory in a radial arm maze task (Naudon *et al*, 2007). Clearly, additional studies are needed to further test the influence of different classes of antidepressants in a standardized manner on behavioral models of learning and memory to fully characterize the influence of these agents.

Antidepressant Treatment Enhances Structural Plasticity and Neurogenesis

Increasing evidence suggests that, in addition to enhancing functional neuroplasticity, antidepressants produce structural plasticity. This is observed at several different levels, including numbers of synapses, spines, dendrites, and even numbers of cells.

At the smallest scale, antidepressant treatment has been reported to increase the number of synapses, determined by electron microscopy (Hajszan *et al*, 2005). This study examined the influence of fluoxetine administration of different durations in ovariectomized rats. As few as 5 days of fluoxetine treatment increases the number of synapses in the CA1 pyramidal cell layer, while longer treatment (14 days) also produced a similar increase in the CA3 pyramidal cell layer. Another study found that chronic administration of a tricyclic antidepressant (amitriptyline) blocked the decrease in spine density in dentate gyrus, CA3 and CA1 cell layers that results from olfactory bulbectomy, a widely used model of depression-like behaviors (Norrholm and Ouimet, 2001; in this study, amitriptyptyline alone did not produce an increase in spine density). Although still preliminary, these labor-intensive studies provide evidence for altered structural plasticity at the levels of both synapse and spine density.

More studies have also been conducted at the level of dendritic morphology in the hippocampus. As described above, McEwen and colleagues have demonstrated that chronic stress (intermittent daily immobilization for 3 weeks) decreases the number and length of apical dendrites of CA3 pyramidal cells in the hippocampus (Watanabe *et al*, 1992; McEwen, 1999). This reduction in dendritic length and complexity is blocked or reversed by chronic administration of an atypical antidepressant (tianeptine), but not by a typical SSRI (fluoxetine; Magariños *et al*, 1999). Additional studies are needed to further test the influence of different classes of antidepressants on dendrite morphology in the hippocampus, both alone and in conjunction with exposure to stress.

Another level of structural plasticity that has received a great deal of attention is the regulation of new cell birth or neurogenesis in the hippocampus. These studies demonstrate that chronic antidepressant administration increases neurogenesis in the adult hippocampus (Malberg *et al*, 2000; reviewed in Warner-Schmidt and Duman, 2006). The increase in neurogenesis is observed with different classes of antidepressants, including SSRIs, selective norepinephrine reuptake inhibitors (SNRIs), MAOIs, atypical antidepressants, and ECS. The neurogenic action of antidepressants requires chronic treatment (14–21 days), with the exception of ECS, which increases neurogenesis 3 days after a single seizure. Antidepressants increase different aspects of neurogenesis, including the rate of proliferation (Malberg *et al*, 2000) and the survival of newborn neurons (Nakagawa *et al*, 2002a). Activation of the cAMP-CREB cascade increases neurogenesis, implicating this pathway in the neurogenic actions of antidepressants (Nakagawa *et al*, 2002b). BDNF is also required for antidepressant regulation of the survival of newborn neurons, though not the proliferation (Sairanen *et al*, 2005).

Establishing the causal role of neurogenesis in the behavioral actions of antidepressants has been challenging, as it is technically difficult to functionally perturb the new neurons without simultaneously disrupting other aspects of behavior or hippocampal function. Evidence to date, however, suggests that the induction of neurogenesis is required for antidepressant action. A casual role for neurogenesis in behavioral change requires that a response to antidepressants be dependent on chronic treatment, because it takes several weeks for newborn neurons to differentiate and mature into functional neurons; such a delayed mechanism of action is consistent with the typical delayed therapeutic response to antidepressants. Two paradigms that are dependent on long-term antidepressant administration are novelty-suppressed feeding and chronic unpredictable stress. These models differ from other standard antidepressant screening paradigms, such as the forced swim and tail suspension tests that respond to acute or short-term antidepressant treatments. An initial study found that blockade of cell proliferation by irradiation blocked the actions of antidepressants in novelty-suppressed feeding and chronic unpredictable stress paradigms (Santarelli *et al*, 2003). In addition, 5-HT1A mutant mice, in which SSRI induction of neurogenesis is blocked, show an attenuated antidepressant response (Santarelli *et al*, 2003). Recent studies using another a genetic model to ablate cells that express GFAP, which includes neural precursors, have confirmed these effects (Rene Hen, personal communication).

Effects of Antidepressant Treatment on Glutamatergic Neurotransmission

Just as stress has been found to alter glutamatergic neurotransmission in ways that can contribute to neuronal atrophy and even cell death (Sapolsky, 2000, 2003), antidepressant treatments can modulate glutamate neurotransmission in neuroprotective ways. Indeed, agents that directly modulate neurotransmission hold promise as novel antidepressants (reviewed in Bleakman *et al*, 2007; Pittenger *et al*, 2007; Witkin *et al*, 2007).

Several lines of evidence suggest that established antidepressants can directly modulate glutamatergic neurotransmission. For example, tricyclic antidepressants directly block the NMDA receptor pore at micromolar concentrations (Reynolds and Miller, 1988). Chronic antidepressant treatment alters the conformation of the NMDA receptor, suggesting some form of longer-term compensatory change (reviewed in Paul and Skolnick, 2003). Such actions of antidepressants on the NMDA receptor may represent a mechanism whereby they can influence the mechanisms of neuroplasticity at the synaptic level.

Antidepressant drugs can also affect AMPA receptor trafficking. The investigational antidepressant riluzole, which was developed for the treatment of amyotrophic lateral sclerosis (Miller *et al*, 2007) but has recently been found to hold promise in the treatment of depression (Zarate *et al*, 2004; Sanacora *et al*, 2007), illustrates this point particularly well. Riluzole reversibly attenuates AMPA-mediated synaptic currents in cultured cells (Albo *et al*, 2004). Both rilulzole and the anticonvulsant lamotrigine, which also has antidepressant properties, increase the surface expression of AMPA subunits GluR1 and GluR2, suggesting an effect on the regulated AMPA receptor trafficking that underlies the conversion of silent to active snapses (Du *et al*, 2007). More conventional antidepressants, such as fluoxetine, can regulate the phosphorylation state and thereby the function of AMPA receptors (Svenningsson *et al*, 2002). These effects of antidepressants that target other molecules and pathways have led to the proposal that direct modulators of AMPA function, or AMPAkines, can function as antidepressants; and substantial evidence supports this idea (reviewed in Witkin *et al*, 2007).

A striking recent finding that emphasizes the interaction between antidepressants and glutamatergic neurotransmission is the rapid antidepressant effect of the NMDA antagonist ketamine. This startling effect was first observed in a small study designed to test the psychotomimetic properties of ketamine in depressed patients (Berman *et al*, 2000) and was more recently replicated in a larger study (Zarate *et al*, 2006). As in the case of increased glutamate after chronic stress, this effect is initially counterintuitive: NMDA activation is trophic and enhances neuroplasticity, so NMDA blockade might be expected to worsen cellular damage. Two observations may make sense of this apparent conundrum. As noted above, synaptic and extrasynaptic NMDA receptors have contrasting effects on neuroplasticity and cellular survival. It is likely that subanesthetic doses of

an NMDA antagonist have a preferential effect on extrasynaptic NMDA receptors, both because the higher-affinity NR2B-containing receptors may be more prevalent in the extrasynaptic space and because the ambient glutamate concentration with which the NMDA antagonist must compete is much lower. In addition, *in vivo* evidence from rats suggests that, again at subanesthetic concentrations, ketamine may preferentially block NMDA receptors on GABAergic interneurons and thereby paradoxically increase synaptic glutamate tone, at least in the frontal cortex (Moghaddam *et al*, 1997). This possibility is supported by the observation that NMDA receptor antagonists increase the expression of BDNF, and activity-dependent gene expression (Metsis *et al*, 1993), and increase adult hippocampal neurogenesis (Gould and Cameron, 1997). Appropriately dosed ketamine may therefore simultaneously decrease activation of extrasynaptic NMDA receptors and enhance glutamate tone at synaptic NMDA receptors, producing a net pro-neuroplasticity effect (reviewed in Pittenger *et al*, 2007).

ANTIDEPRESSANTS INCREASE NEUROPLASTICITY-RELATED SIGNALING PATHWAYS

Overlap between the molecular actions of synaptic plasticity and those targeted by antidepressants provided some of the earliest and strongest evidence for mechanistic overlap between the two phenomena. These studies demonstrate that some of the key signaling components identified as critical regulators of LTP and synaptic plasticity are also regulated by antidepressants and are required for the cellular and behavioral actions of antidepressant treatments.

Antidepressants Upregulate the cAMP-PKA-CREB Cascade

Evidence from a number of different studies has demonstrated that antidepressant treatments upregulate the cAMP signal transduction pathway. This includes increased levels of Gαs coupling to adenylyl cyclase in the hippocampus and frontal cortex and resulting elevation of cAMP production in response to chronic administration of different classes of antidepressants (reviewed in Donati and Rasenick, 2003). Levels of PKA are also reported to be upregulated by chronic administration of different classes of antidepressants, including tricyclics, MAOIs, and ECS (Nestler *et al*, 1989; Perez *et al*, 1989). Levels of PKA are increased in particulate fractions, and subcellular fractionation studies demonstrate increases in both the nuclear (Nestler *et al*, 1989; Tiraboschi *et al*, 2004b) and microtubule fractions (Perez *et al*, 2000). The increases in nuclear PKA levels are more consistent with SNRI than SSRI antidepressants, and vary somewhat depending on the brain region examined (PFC *vs* hippocampus) and type of enzyme activity (basal *vs* cAMP-stimulated) in different studies.

Studies demonstrating an increase in nuclear levels of PKA also suggest that the actions of antidepressant treatments could involve the regulation of transcription factors, such as CREB. Early studies provided support for this hypothesis, demonstrating that chronic administration of different classes of antidepressants, including SSRI and SNRI agents, MAOIs, and ECS, increases the expression and function of CREB in the PFC and hippocampus (Nibuya *et al*, 1996; Frechilla *et al*, 1998; for reviews see Tardito *et al*, 2006; Blendy, 2006). These studies demonstrate that chronic, but not acute, antidepressant treatment increases levels of CREB mRNA and immunoreactivity, as well as levels of CRE binding in the hippocampus. Subsequent studies have also demonstrated that the phosphorylation and transcriptional activity of CREB is increased by chronic antidepressant treatment (Thome *et al*, 2000; Tiraboschi *et al*, 2004b). Analysis of CREB and nuclear localization of several protein kinases in the same study indicates that PKA is less likely to account for the induction of CREB phosphorylation than CAMKIV and MAPK, although a role for PKA cannot be completely excluded (Tiraboschi *et al*, 2004b).

Regulation of CREB is likely to be important in antidepressant responses in experimental animals. Viral expression of CREB in the hippocampus produces antidepressant effects in behavioral models of despair and antidepressant response, including the forced swim test and learned helplessness model (Chen *et al*, 2001a, b). Studies of mutant mice have reported that CREB hypomorphic mice have normal antidepressant responses in the FST (Conti *et al*, 2002); this discrepancy may result, in part, from adaptive changes in these knockout mice from partial CREB knockout throughout the brain and throughout development. Drugs that more directly activate the cAMP-CREB cascade, such as the cAMP-specific phosphodiesterase type IV (PDE4) inhibitor rolipram, produce antidepressant effects in behavioral models (reviewed in O'Donnell and Zhang, 2004). The cAMP-CREB cascade also increases adult hippocampal neurogenesis (Nakagawa *et al*, 2002a, b), which, as described above, has been implicated in the actions of antidepressant treatment (Santarelli *et al*, 2003; Duman, 2004).

In humans, there may be a link between genetic polymorphisms in the CREB gene and major depression (Zubenko *et al*, 2003). A recent study also suggested an association between a CREB polymorphism and anger expression (which may relate to suicide risk) in men with depression (Perlis *et al*, 2007). Post-mortem studies suggest that reduced CREB function may contribute to clinical depression, and that its upregulation may also be an important component of antidepressant response in humans. CREB expression in temporal cortex is decreased in depressed suicide subjects (Dowlatshahi *et al*, 1998; Odagaki *et al*, 2001; Dwivedi *et al*, 2003); CREB phosphorylation has also been reported to be reduced in post-mortem tissue (Yamada *et al*, 2003). In addition, CREB is upregulated by pre-mortem antidepressant treatment in post-mortem tissue (Dowlatshahi *et al*, 1998), as well as in blood platelets (Koch *et al*, 2002). While post-mortem studies such as these are particularly challenging, they provide critical evidence that mechanisms elucidated in experimental animals, such as the role of CREB in the antidepressant response, are of clinical relevance in human patients.

These studies provide convincing evidence that the cAMP-PKA-CREB cascade is regulated by antidepressant treatment, and that activation of this pathway can produce

an antidepressant response in behavioral models. The evidence linking these effects to synaptic plasticity is less robust. As described above, the cAMP–PKA–CREB pathway has a central role in many forms of neuroplasticity, and antidepressants can enhance LTP and memory; but this argument for a role for CREB-mediated neuroplasticity is circumstantial. No studies to date have directly tested the role of this pathway in the actions of antidepressants, and future studies will be required to test this hypothesis.

Antidepressants Regulate CaMKII

A number of studies have investigated the regulation of CaMKII by antidepressants. Several studies have reported that chronic, but not acute, antidepressant treatment increases CaMKII enzymatic activity in neuronal cell bodies of the hippocampus, due to an increase in autophosphorylation of CaMKII at Thr^{286} (Popoli *et al*, 1995; Celano *et al*, 2003; Tiraboschi *et al*, 2004a). As described above, phosphorylation at this site leads to Ca^{2+}-independent activation of CaMKII and therefore extended enzymatic activity. Subsequent studies demonstrate that there is also an increase in CAMKII expression, enzymatic activity, and phosphorylation in synaptic vesicles of the hippocampus and PFC after antidepressant treatment (Celano *et al*, 2003), but a decrease in phosphorylation of CaMKII in synaptic terminals and synaptic membranes (Bonanno *et al*, 2005). This apparent contradiction has been resolved by a recent study from the same group, demonstrating that antidepressant treatments (with either an SSRI or an SNRI) cause a redistribution of CaMKII from synaptic membranes to vesicles and a corresponding decrease in membrane-associated synapsin I (Barbiero *et al*, 2007). Since vesicles localized with the synaptic membranes are closely related to the readily releasable pool of glutamate, this redistribution suggests that antidepressant treatment reduces the potential for glutamate release. This is supported by a functional study demonstrating that depolarization-induced release of glutamate is decreased (Bonanno *et al*, 2005).

If this interpretation is correct—that modulation of CaMKII by antidepressants leads to a net decrease in synaptic release of glutamate—antidepressant treatment would decrease the potential for synaptic plasticity, at least via this mechanism. While a reduction in membrane-associated synaptic vesicles could decrease the overall release of glutamate, the resulting effect could be to increase the signal-to-noise ratio such that only salient stimuli lead to synaptic plasticity-related changes in the hippocampus and PFC. An enhanced signal-to-noise ratio may be further enhanced by potentiation of other mechanisms contributing to plasticity through alterations in downstream postsynaptic mechanisms such as cAMP-mediated and MAPK signaling (as reviewed above and below, respectively). Alternatively, the reported upregulation of CAMKII in cell bodies, as opposed to presynaptic terminals, could lead to other types of synaptic plasticity and neuroprotection that also contribute to the actions of antidepressant treatments.

Antidepressants Upregulate the MAPK Cascade

As reviewed above, the MAPK cascade has also been implicated in some forms of long-lasting neuroplasticity. The MAPK pathway is also activated by BDNF and other neurotrophic factors. A number of reports suggest that antidepressants influence the expression or phosphorylation of the kinases in this pathway as well. One study found that chronic administration of different classes of antidepressants increases levels of ERK1 and ERK2 in total homogenates of hippocampus and PFC, although levels of phospho-ERK were not altered (Tiraboschi *et al*, 2004b). This study also reported increases in levels of phospho-ERK1/2 in nuclear fractions, under some conditions, suggesting that ERK signaling may contribute to increased phosphorylation and activation of CREB. However, the situation may be more complicated than this; another study found that fluoxetine administration decreased levels of phospho-ERK1/2 in nuclear and/or cytosolic fractions of hippocampus and PFC, while imipramine increased levels of phospho-ERK in PFC only (Fumagalli *et al*, 2005). Interestingly, chronic administration of an atypical antipsychotic (olanzapine), which has antidepressant efficacy and increases the actions of SSRIs in clinical studies (Nemeroff, 2006), increases levels of ERK1 and ERK2 in PFC fractions (Fumagalli *et al*, 2006).

As in the case of the cAMP/CREB pathway, other lines of evidence suggest that MAPK modulation by antidepressant treatment is likely to play a causal role in the antidepressant response. Agents that inhibit the MAPK pathway increase behavioral despair and helplessness in the forced swim test and learned helplessness paradigms and block the effects of antidepressants in rodents (Duman *et al*, 2007). An earlier study found contrasting effects of MAPK inhibitors, but these effects may be explained by the acute locomotor-activating actions of these agents (Einat *et al*, 2003). Finally, post-mortem studies report that ERK activity and expression are decreased in depressed suicide subjects (Dwivedi *et al*, 2001). More recent studies have reported that ERK5 is also decreased in suicide subjects (Dwivedi *et al*, 2007) and that the upstream kinases responsible for activation of ERK are decreased (Dwivedi *et al*, 2006).

Taken together these studies indicate that antidepressants upregulate the MAPK pathway, opposing the actions of suicide/depression, and that blockade of the MAPK cascade results in despair/helplessness and inhibition of the antidepressant actions in behavioral models. However, there are also conflicting reports on antidepressant regulation of phospho-ERK1/2, depending on the brain region, subcellular fraction, and antidepressant tested.

ANTIDEPRESSANTS UPREGULATE NEUROTROPHIC FACTOR EXPRESSION

Another line of evidence supporting the hypothesis that antidepressant treatment leads to altered neuroplasticity is the regulation of neurotrophic factors by antidepressants. As described above, certain neurotrophic factors, such as BDNF, are regulated by activity; BDNF has been shown to be necessary and sufficient for induction of LTP and performance in some models of learning and memory (reviewed in Pang and Lu, 2004). Furthermore, BDNF is critical to numerous forms of neuroplasticity, and is reduced by chronic stress, as reviewed above. We now

review the evidence that antidepressant treatment increases BDNF and other trophic factors.

Antidepressant Treatment Increases the Expression of BDNF

Chronic antidepressant administration increases the expression of BDNF in the hippocampus and PFC (Nibuya *et al*, 1995; reviewed in Duman and Monteggia, 2006); this effect contrasts with the inhibition of BDNF by chronic stress, as reviewed above. Increased expression of BDNF is observed with different classes of antidepressants, including SSRI, SNRI, MAOI, atypical agents, and ECS. The induction of BDNF is dependent on chronic treatment for 2–3 weeks and is pharmacologically specific: other types of psychotropic drugs do not increase BDNF in the hippocampus and PFC.

There have been a number of conflicting studies, reporting no change or even decreased BDNF expression. These discrepancies are most likely due to differences in the time course or dose of antidepressant tested, or the time after the last treatment (reviewed in Duman and Monteggia, 2006). Since BDNF expression is also decreased by stress and increased by neuronal activity, the individual experimental conditions (eg housing, treatment) are critical for the outcome of these studies.

Roles for BDNF in depression and antidepressant responses are supported by several additional lines of evidence. First, levels of BDNF are decreased in the brains of depressed patients and increased in patients receiving antidepressant treatment at the time of death (Chen *et al*, 2001b; Dwivedi *et al*, 2003; Karege *et al*, 2005). Surprisingly, levels of BDNF in serum of depressed subjects are also decreased; this decrease is reversed by antidepressant treatment, suggesting that it may prove to be a useful, relatively noninvasive biomarker of the depressed state (Karege *et al*, 2002; Shimizu *et al*, 2003; reviewed in Duman and Monteggia, 2006).

Second, BDNF infusions produce an antidepressant response in behavioral models, including the forced swim and learned helplessness paradigms (Siuciak *et al*, 1997; Shirayama *et al*, 2002; Hoshaw *et al*, 2005); and the response to antidepressant treatment is blocked in mice with forebrain-specific impairment of either BDNF or its receptor, TrkB (Saarelainen *et al*, 2003; Monteggia *et al*, 2004). Conditional BDNF mutant mice also display an increase in depression-like behavior in the forced swim and sucrose preference tests, although these effects are specific to female mice (Monteggia *et al*, 2007).

Third, a recent clinical study has reported that a BDNF polymorphism (val66met), when combined with the S allele of the 5-HT transporter, increases the risk for depression in children exposed to trauma (Kaufman *et al*, 2006). The met allele decreases the processing and release of BDNF and has been linked to reduced episodic memory and hippocampal function in humans (Egan *et al*, 2003). The same allele produces a striking anxiety phenotype when introduced into mice (Chen *et al*, 2006).

Together, these studies provide strong support for the hypothesis that reduced BDNF expression contributes to depression-like behavior in animal models and may lead to similar effects in humans, and that antidepressant treatment increases or reverses this deficit. These findings are consistent with the hypothesis that BDNF-induced neuroplasticity contributes to the actions of antidepressants, although it is also possible that other BDNF actions (such as neuroprotection) also play a role.

Antidepressant Treatment Increases the Expression of VEGF

Recent studies have provided evidence that another trophic factor, VEGF, contributes to the actions of stress and antidepressant treatments. Previous studies have demonstrated that VEGF is regulated by neuronal activity and that VEGF increases hippocampal LTP and synaptic transmission (Cao *et al*, 2004). We have recently reported that chronic antidepressant administration, including SSRI, SNRI, and ECS, increases the expression of VEGF mRNA and protein in the hippocampus (Newton *et al*, 2003; Warner-Schmidt and Duman, 2007). As is the case with BDNF, this increase is opposite to the stress-induced decrease in VEGF expression in the hippocampus (Heine *et al*, 2005). VEGF infusions into the lateral ventricles increase neurogenesis and produce antidepressant responses in four different behavioral models, including those responsive to short- (forced swim and learned helplessness) and long-term antidepressant treatments (sucrose preference and novelty-induced hypophagia). In contrast, blockade of VEGF signaling through one of its receptors, VEGF-R2 (also known as Flk-1), blocked the behavioral and neurogenic effects of antidepressants in all four behavioral models (Warner-Schmidt and Duman, 2007).

The results of these studies suggest that VEGF is necessary and sufficient for the behavioral and neurogenic actions of antidepressants. Moreover, the results demonstrate that VEGF contributes to the actions of antidepressants in both short- and long-term responses. It is possible that the rapid effects of VEGF and antidepressants are mediated by regulation of neuroplasticity and synaptic transmission, while the more long-term effects occur via regulation of neurogenesis. Further studies will be required to more completely delineate the cellular mechanisms underlying the actions and interactions of VEGF and antidepressants.

It is striking that multiple neurotrophins have now been implicated in antidepressant action. Indeed, the list may not end here—there is evidence that insulin-like growth factor-1 and fibroblast growth factor-2 are involved in stress, depression, and antidepressant responses (Duman and Monteggia, 2006). There are several possible explanations for this apparent redundancy. First, multiple trophic factors may provide redundancy for neuroplasticity and neuroprotection—critical neuronal functions. However, complete redundancy of the different factors is not supported by data, suggesting that BDNF and VEGF are individually necessary, not just sufficient, for antidepressant responses. It is possible that the different trophic factors exert different, complementary actions, in addition to functions that have some overlap. Additional studies will be needed to fully characterize the interactions of these neurotrophic/growth factor systems.

LIMITATIONS OF THE NEUROPLASTICITY HYPOTHESIS

The neuroplasticity hypothesis suggests that impaired mechanisms of neuroplasticity are a core pathophysiological feature of MDD, that chronic stress is an important causal factor in the development of this impairment, and that antidepressant treatments act, at least in part, through mitigation of impaired mechanisms of plasticity. This hypothesis unifies much clinical and preclinical data, which we have sought to summarize above. However, the hypothesis also has important limitations.

One that has received attention recently is that alterations of neuroplasticity in certain brain regions produce a pro-depressive effect. For example, expression of CREB in the nucleus accumbens increases behavioral despair and helplessness in the forced swim and learned helpless paradigms, and CREB inhibition has an antidepressant effect (Pliakas *et al*, 2001; Newton *et al*, 2002). Similarly, increased expression of BDNF in the mesolimbic dopamine system produces a pro-depressive effect in the forced swim test (Eisch *et al*, 2003) and social defeat stress models of depression (Berton *et al*, 2006). These studies indicate that neuroplasticity in the mesolimbic dopamine circuit can produce effects that are opposite to the antidepressant effects observed in the hippocampus.

Neuroplasticity in the amygdala may likewise work contrary to antidepressant effects, though here the argument is more circumstantial. Chronic stress and glucocorticoid treatment increase dendritic branching and synaptic connectivity in the amygdala, perhaps paralleling the increased size of the amygdala reported in some neuroimaging studies of depressed patients. It seems likely that these plastic events contribute to the development of a depressive state, rather than contributing to antidepressant response.

This suggestion raises the concerning possibility that novel therapeutic strategies aimed at the enhancement of neuroplasticity may have unintended pro-depressant effects through their impact on the nucleus accumbens or other structures outside the hippocampal system that has informed most of our discussion. Early studies suggest that the predominant effect of activating neuroplasticity-associated signaling systems is an antidepressant response. Systemic administration of rolipram, which activates the cAMP-CREB cascade, produces an antidepressant effect in behavioral models (O'Donnell and Zhang, 2004). Similarly, intraventricular infusions of BDNF produce an antidepressant response in the forced swim test (Hoshaw *et al*, 2005). Despite these promising results, further studies in a variety of behavioral models are needed to fully characterize the global effects of activation of neuroplasticity-associated pathways.

RELEVANCE OF ENHANCEMENT IN NEUROPLASTICITY TO BEHAVIORAL MODELS OF DEPRESSION

Although enhanced neuroplasticity is considered a desirable adaptation, the relationship with the actions of antidepressants is not readily interpretable. Are the enhancements in neuroplasticity we have reviewed part of an important pathway that contributes to antidepressant effects? Are they a consequence of altered mood state or homeostasis of modulatory neurotransmitters or hormone systems? Or is the association of enhanced neuroplasticity with antidepressant effect an epiphenomenon—suggestive of shared underlying mechanisms, perhaps, but not necessarily indicative of a causal relationship between the two phenomena?

The nature of the behaviors in rodents that are widely used to model aspects of depression and have proven validity for predicting antidepressant response may shed some light on this question. Most of the behaviors used to assay antidepressant effect in rodents—learned helplessness, the forced swim test, and the tail suspension test, for example—are models of behavioral despair and coping. Animals are placed in an adverse circumstance from which it is difficult or impossible to escape, and they enter a behavioral state of passivity in which they no longer attempt to do so. It may be that enhancements in neuroplasticity, and the corresponding increased capacity to adapt and learn, lead to an enhanced potential repertoire of behaviors or capacity to explore new escape options in these adverse circumstances, and thus reduces the tendency to enter such a state of behavioral despair. This interpretation predicts that enhanced neuroplasticity is, indeed, causally important for reduced depression-like behaviors, at least in models of behavioral despair. It is to be hoped that as our understanding of the complex relationships between stress, depression, and neuroplasticity grows, the functional interrelationship of one to another will become more clear. Even at this early point, the extensive mechanistic parallels between these disparate phenomena suggest a deep causal connection.

SUMMARY AND CONCLUSIONS

The molecular mechanisms of antidepressant response overlap with those of neuroplasticity to a striking degree. We have described three levels of neuroplasticity in which this parallel can be seen. Synaptic plasticity (including morphological change in dendritic spines and other concomitants of synaptic connections) has been well characterized in many brain regions, most prominently the hippocampus. Many of the molecular mechanisms of synaptic plasticity are induced by antidepressants; again, the hippocampus presents the best-characterized example of this fact. Parallels are seen both with the mechanisms of short-term plasticity—such as the insertion of AMPA receptors into the membrane and the modulation of CaMKII—and with those of long-term plasticity—such as the activation of transcription factors such as CREB and the induction of plasticity-enhancing genes such as BDNF.

Parallels are also seen at the level of dendritic morphology. Increases in dendritic branching can be induced by enriched environments (van Praag *et al*, 2000), growth factors, and other neuroplasticity-inducing manipulations. Similar increases in dendritic length and complexity can be induced by antidepressant treatment. Finally, neurogenesis represents a more recently appreciated but profound mechanism of neuroplasticity, permitting the insertion of new neurons into the dentate gyrus of the adult animal. Here, too, neurogenesis is stimulated both by plasticity-inducing stimuli (such as environmental enrichment, exercise, and electrical stimulation) and by antidepressant treatments.

While it is difficult to examine these mechanisms in tissue from depressed humans, evidence exists for impairment of neuroplasticity in major depression. Molecular concomitants of certain aspects of synaptic plasticity—such as expression and phosphorylation of the transcription factor CREB—are reduced in post-mortem tissue from depressed subjects. Depression leads to shortening and reduced complexity of dendritic trees, in the hippocampus and the PFC. It remains to be determined whether neurogenesis is reduced in patients with major depression (the one study published to date was negative; Reif *et al*, 2006); but neurogenesis has been shown to occur in the human hippocampus in adulthood (Eriksson *et al*, 1998), and thorough examination of whether it is reduced in major depression is certainly warranted by the strength of the preclinical evidence.

Stress is well established to contribute to the development of major depression in susceptible individuals. It is likely that any comprehensive account of the pathophysiology of mood disorders will include an important causal role for psychological and physiological stress. It is therefore striking that chronic stress, too, opposes many forms of neuroplasticity. As reviewed above, chronic stress reduces synaptic plasticity and dendritic spines, reduces the length and complexity of dendrites, and impairs neurogenesis.

The convergence of these literatures—neuroplasticity, antidepressant response, and the consequences of chronic stress—on an overlapping set of molecular and cellular mechanisms is increasingly apparent and suggests a deep connection between these three phenomena. Further explorations and better understanding of these relationships is likely to point the way toward a deeper understanding of affective disorders and, it is to be hoped, toward novel treatments. This will include the development and use of novel strategies for imaging structural and molecular pathways in human subjects *in vivo*, as well as post-mortem and genetic studies to identify and define the exact alterations underlying the disruption of neuroplasticity in depressed subjects. The findings demonstrating altered neuroplasticity have already provided novel targets for drug development and future studies will continue to identify and refine the information for further development of more effective therapeutic interventions.

DISCLOSURE/CONFLICT OF INTEREST

RSD is a consultant for Taisho Pharmaceuticals and Psychogenics, has received funding for investigator initiated research from Lilly, Organon, Pfizer, Sepracor, and Roche, and has received speaker honorariums from Lilly, Pfizer, Sepracor, Roche, Bristol-Myers-Squibb and Wyeth-Ayerst. CP has no disclosures or conflicts to report.

REFERENCES

Abel T, Martin KC, Bartsch D, Kandel ER (1998). Memory suppressor genes: inhibitory constraints on the storage of long-term memory. *Science* **279**: 338–341.

Abel T, Nguyen PV, Barad M, Deuel TA, Kandel ER, Bourtchouladze R (1997). Genetic demonstration of a role for PKA in the late phase of LTP in hippocampus-based long-term memory. *Cell* **88**: 615–626.

Abercrombie HC, Schaefer SM, Larson CL, Oakes TR, Lindgren KA, Holden JE *et al* (1998). Metabolic rate in the right amygdala predicts negative affect in depressed patients. *Neuroreport* **9**: 3301–3307.

Ahmed T, Frey JU, Korz V (2006). Long-term effects of brief acute stress on cellular signaling and hippocampal LTP. *J Neurosci* **26**: 3951–3958.

Albo F, Pieri M, Zona C (2004). Modulation of AMPA receptors in spinal motor neurons by the neuroprotective agent riluzole. *J Neurosci Res* **78**: 200–207.

Alfonso J, Aguero F, Sanchez DO, Flugge G, Fuchs E, Frasch AC *et al* (2004). Gene expression analysis in the hippocampal formation of tree shrews chronically treated with cortisol. *J Neurosci Res* **78**: 702–710.

Alfonso J, Frick LR, Silberman DM, Palumbo ML, Genaro AM, Frasch AC (2006). Regulation of hippocampal gene expression is conserved in two species subjected to different stressors and antidepressant treatments. *Biol Psych* **59**: 244–251.

Allain H, Lieury A, Brunet-Bourgin F, Mirabaud C, Trebon P, Le Coz F *et al* (1992). Antidepressants and cognition: comparative effects of moclobemide, viloxazine and maprotiline. *Psychopharmacology (Berl)* **106**: S56–S61.

Alonso G (2000). Prolonged corticosterone treatment of adult rats inhibits the proliferation of oligodendrocyte progenitors present throughout white and gray matter regions of the brain. *Glia* **31**: 219–231.

Altman J, Das GD (1965). Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. *J Comp Neurol* **124**: 319–335.

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders (DSM) IV-TR*. American Psychiatric Press: Washington, DC.

Atkins CM, Selcher JC, Petraitis JJ, Trzaskos JM, Sweatt JD (1998). The MAP cascade is required for mammalian associative learning. *Nat Neurosci* **1**: 602–609.

Autry AE, Grillo CA, Piroli GG, Rothstein JD, McEwen BS, Reagan LP (2006). Glucocorticoid regulation of GLT-1 glutamate transporter isoform expression in the rat hippocampus. *Neuroendocrinology* **83**: 371–379.

Bagley J, Moghaddam B (1997). Temporal dynamics of glutamate efflux in the prefrontal cortex and in the hippocampus following repeated stress: effects of pretreatment with saline or diazepam. *Neuroscience* **77**: 65–73.

Banasr M, Valentine GW, Li XY, Gourley S, Taylor J, Duman RS (2007). Chronic stress decreases cell proliferation in adult cerebral cortex of rat: reversal by antidepressant treatment. *Biol Psych* **62**: 496–504.

Barad M, Bourtchouladze R, Winder DG, Golan H, Kandel ER (1998). Rolipram, a type IV-specific phosphodiesterase inhibitor, facilitates the establishment of long-lasting long-term potentiation and improves memory. *Proc Natl Acad Sci USA* **95**: 15020–15025.

Barbiero VS, Giambelli R, Musazzi L, Tiraboschi E, Tardito D, Perez J *et al* (2007). Chronic antidepressants induce redistribution and differential activation of alpha CaM Kinase II between presynaptic compartments. *Neuropsychopharmacology*; epub ahead of print.

Bartoletti A, Cancedda L, Reid SW, Tessarollo L, Porciatti V, Pizzorusso T *et al* (2002). Heterozygous knock-out mice for brain-derived neurotrophic factor show a pathway-specific impairment of long-term potentiation but normal critical period for monocular deprivation. *J Neurosci* **22**: 10072–10077.

Bartsch D, Casadio A, Karl KA, Serodio P, Kandel ER (1998). CREB1 encodes a nuclear activator, a repressor, and a cytoplasmic modulator that form a regulatory unit critical for long-term facilitation. *Cell* **95**: 211–223.

Baxter Jr LR, Schwartz JM, Phelps ME, Mazziotta JC, Guze BH, Selin CE *et al* (1989). Reduction of prefrontal cortex glucose

metabolism common to three types of depression. *Arch Gen Psych* **46**: 243–250.
Berman RM, Cappiello A, Anand A, Oren DA, Heninger GR, Charney DS *et al* (2000). Antidepressant effects of ketamine in depressed patients. *Biol Psych* **47**: 351–354.
Berton O, McClung CA, Dileone RJ, Krishnan V, Renthal W, Russo SJ *et al* (2006). Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress. *Science* **311**: 868–878.
Bito H, Deisseroth K, Tsien RW (1996). CREB phosphorylation and dephosphorylation: a Ca^{2+} and stimulus duration-dependent switch for hippocampal gene expression. *Cell* **87**: 1203–1214.
Bleakman D, Alt A, Witkin JM (2007). AMPA receptors in the therapeutic management of depression. *CNS Neurol Disord—Drug Targets* **6**: 117–126.
Blendy JA (2006). The role of CREB in depression and antidepressant treatment. *Biol Psych* **59**: 1144–1150.
Bodnoff S, Humphreys A, Lehman J, Diamond D, Rose G, Meaney M (1995). Enduring effects of chronic corticosterone treatment on spatial learning, synaptic plasticity and hippocampal neuropathology in young and mid-aged rats. *J Neurosci* **15**: 61–69.
Bonanno G, Giambelli R, Raiteri L, Tiraboschi E, Zappettini S, Muszzi L *et al* (2005). Chronic antidepressants reduce depolarization-evoked glutamate release and protein interactions favoring formation of SNARE complex in hippocampus. *J Neurosci* **25**: 3270–3279.
Bremner JD, Narayan M, Anderson ER, Staib LH, Miller HL, Charney DS (2000). Hippocampal volume reductions in major depression. *Am J Psych* **157**: 115–117.
Bukalo O, Fentrop N, Lee AY, Salmen B, Law JW, Wojak CT *et al* (2004). Conditional ablation of the neural cell adhesion molecule reduces precision of spatial learning, long-term potentiation, and depression in the CA1 subfield of the mouse hippocampus. *J Neurosci* **24**: 1565–1577.
Cao L, Jiao X, Zuzga DS, Liu Y, Fong DM, Young D *et al* (2004). VEGF links hippocampal activity with neurogenesis, learning and memory. *Nat Genet* **36**: 827–835.
Carlezon Jr WA, Duman RS, Nestler EJ (2005). The many faces of CREB. *Trends Neurosci* **28**: 436–445.
Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H *et al* (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* **301**: 386–389.
Celano E, Tiraboschi E, Consogno E, D'Irso G, Mbakop MP, Gennarelli M *et al* (2003). Selective regulation of presynaptic calcium/calmodulin-dependent protein kinase II by psychotropic drugs. *Biol Psych* **53**: 442–449.
Chambers RA, Potenza MN, Hoffman RE, Miranker W (2004). Simulated apoptosis/neurogenesis regulates learning and memory capabilities of adaptive neural networks. *Neuropsychopharmacology* **29**: 747–758.
Chen AC, Shirayama Y, Shin KH, Neve RL, Duman RS (2001a). Expression of the cAMP response element binding protein (CREB) in hippocampus produces an antidepressant effect. *Biol Psych* **49**: 753–762.
Chen B, Dowlatshahi D, MacQueen GM, Wang JF, Young LT (2001b). Increased hippocampal BDNF immunoreactivity in subjects treated with antidepressant medication. *Biol Psych* **50**: 260–265.
Chen ZY, Jing D, Bath KG, Ieraci A, Khan T, Siao CJ *et al* (2006). Genetic variant BDNF (Val66Met) polymorphism alters anxiety-related behavior. *Science* **314**: 140–143.
Conrad CD, Galea LA, Kuroda Y, McEwen BS (1996). Chronic stress impairs rat spatial memory on the Y maze, and this effect is blocked by tianeptine pretreatment. *Behav Neurosci* **110**: 1321–1334.
Conrad CD, LeDoux JE, Magannos AM, McEwen BS (1999). Repeated restraint stress facilitates fear conditioning independently of causing hippocampal CA3 dendritic atrophy. *Behav Neurosci* **113**: 902–913.
Conti AC, Cryan JF, Dalvi A, Lucki I, Blendy JA (2002). cAMP response element-binding protein is essential for the upregulation of brain-derived neurotrophic factor transcription, but not the behavioral or endocrine responses to antidepressant drugs. *J Neurosci* **22**: 3262–3268.
Cook SC, Wellman CL (2004). Chronic stress alters dendritic morphology in rat medial prefrontal cortex. *J Neurobiol* **60**: 236–248.
Cotter D, Mackay D, Chana G, Beasley C, Landau S, Everall IP (2002). Reduced neuronal size and glial cell density in area 9 of the dorsolateral prefrontal cortex in subjects with major depressive disorder. *Cereb Cortex* **12**: 386–394.
Cotter D, Mackay D, Landau S, Kerwin R, Everall I (2001). Reduced glial cell density and neuronal size in the anterior cingulate cortex in major depressive disorder. *Arch Gen Psych* **58**: 545–553.
De Kloet ER (2004). Hormones and the stressed brain. *Ann NY Acad Sci* **1018**: 1–15.
De Quervain D, Roozendaal B, McGaugh J (1998). Stress and glucocorticoids impair retrieval of long-term spatial memory. *Nature* **394**: 787–790.
De Quervain D, Roozendaal B, Nitsch R, McGaugh J, Hock C (2000). Acute cortisone administration impairs retrieval of long-term declarative memory in humans. *Nat Neurosci* **3**: 313–317.
Deisseroth K, Singla S, Toda H, Monje M, Palmer TD, Malenka Rc (2004). Excitation-neurogenesis coupling in adult neural stem/progenitor cells. *Neuron* **42**: 535–552.
Di Cristo G, Berardi N, Cancedda L, Pizzorusso T, Putingnano E, Ratto GM *et al* (2001). Requirement of ERK activation for visual cortical plasticity. *Science* **292**: 2337–2340.
Diamond D, Park C, Heman K, Rose G (1999). Exposing rats to a predator impairs spatial working memory in the radial arm water maze. *Hippocampus* **9**: 542–552.
Donati RJ, Rasenick MM (2003). G protein signaling and the molecular basis of antidepressant action. *Life Sci* **73**: 1–17.
Dowlatshahi D, MacQueen GM, Wang JF, Young LT (1998). Increased temporal cortex CREB concentrations and antidepressant treatment in major depression. *Lancet* **352**: 1754–1755.
Dranovsky A, Hen R (2006). Hippocampal neurogenesis: regulation by stress and antidepressants. *Biol Psych* **59**: 1136–1143.
Drapeau E, Mayo W, Aurousseau C, Le Moal M, Piazza P, Abrous DN (2003). Spatial memory performances of aged rats in the water maze predict levels of hippocampal neurogenesis. *Proc Natl Acad Sci USA* **100**: 14385–14390.
Drevets WC, Videen TO, Price JL, Preskorn SH, Carmichael ST, Raichle ME (1992). A functional anatomical study of unipolar depression. *J Neurosci* **12**: 3628–3641.
Drevets WC (2003). Neuroimaging abnormalities in the amygdala in mood disorders. *Ann NY Acad Sci* **985**: 420–444.
Du J, Suzuki K, Wei Y, Wang Y, Blumenthal R, Chen Z *et al* (2007). The anticonvulsants lamotrigine, riluzole, and valproate differentially regulate AMPA receptor membrane localization: relationship to clinical effects in mood disorders. *Neuropsychopharmacology* **32**: 793–802.
Duman CH, Schlesinger L, Kodama M, Russell DS, Duman RS (2007). A role for MAPK signaling in behavioral models of depression and antidepressant treatment. *Biol Psych* **61**: 661–670.
Duman RS, Heninger GR, Nestler EJ (1997). A molecular and cellular theory of depression. *Arch Gen Psych* **54**: 597–606.
Duman RS, Monteggia LM (2006). A neurotrophic model for stress-related mood disorders. *Biol Psych* **59**: 1116–1127.
Duman RS (2004). Depression: a case of neuronal life and death? *Biol Psych* **56**: 140–145.
Dunn RT, Kimbrell TA, Ketter TA, Frye MA, Willis MW, Luckenbaugh DA *et al* (2002). Principal components of the Beck Depression Inventory and regional cerebral metabolism in unipolar and bipolar depression. *Biol Psych* **51**: 387–399.
Dwivedi Y, Rizavi HS, Roberts RC, Conley RC, Tamminga CA, Pandey GN (2001). Reduced activation and expression of ERK

1/2 MAP kinase in the post-mortem brain of depressed suicide subjects. *J Neurochem* **77**: 916–928.

Dwivedi Y, Rao JS, Rizavi HS, Kotowski J, Conley RR, Roberts RC *et al* (2003). Abnormal expression and functional characteristics of cyclic adenosine monophosphate response element binding protein in postmortem brain of suicide subjects. *Arch Gen Psych* **60**: 273–282.

Dwivedi Y, Rizavi HS, Conley RR, Pandey GN (2006). ERK MAP kinase signaling in post-mortem brain of suicide subjects: differential regulation of upstream Raf kinases Raf-1 and B-Raf. *Mol Psych* **11**: 86–98.

Dwivedi Y, Rizavi HS, Teppen T, Sasaki N, Chen H, Zhang H *et al* (2007). Aberrant extracellular signal-regulated kinase (ERK) 5 signaling in hippocampus of suicide subjects. *Neuropsychopharmacology*; epub ahead of print.

Egan MF, Kojima M, Callicot JH, Goldberg TE, Kolachana BS, Bertolino A *et al* (2003). The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell* **112**: 257–269.

Einat H, Yuan P, Gould TD, Li J, Du J, Zhang L *et al* (2003). The role of the extracellular signal-regulated kinase signaling pathway in mood modulation. *J Neurosci* **23**: 7311–7316.

Eisch AJ, Bolanos CA, de Wit J, Simonak RD, Pudiak CM, Barrot M *et al* (2003). Brain-derived neurotrophic factor in the ventral midbrain-nucleus accumbens pathway: a role in depression. *Biol Psych* **54**: 994–1005.

Engert F, Bonhoeffer T (1999). Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature* **399**: 66–70.

Epstein J, Pan H, Kocsis JH, Yang Y, Butler T, Chusid J *et al* (2006). Lack of ventral striatal response to positive stimuli in depressed versus normal subjects. *Am J Psych* **163**: 1784–1790.

Eriksson PS, Perfilieva E, Björk-Eriksson T, Alborm AM, Nordborg C, Peterson DA *et al* (1998). Neurogenesis in the adult human hippocampus. *Nat Med* **4**: 1313–1317.

Focking M, Holker I, Trapp T (2003). Chronic glucocorticoid receptor activation impairs CREB transcriptional activity in clonal neurons. *Biochem Biophys Res Comm* **304**: 720–723.

Foy MR, Stanton ME, Levine S, Thompson RF (1987). Behavioral stress impairs long-term potentiation in rodent hippocampus. *Behav Neural Biol* **48**: 138–149.

Franklin TB, Perrot-Sinal TS (2006). Sex and ovarian steroids modulate brain-derived neurotrophic factor (BDNF) protein levels in rat hippocampus under stressful and non-stressful conditions. *Psychoneuroendocrinology* **31**: 38–48.

Frechilla D, Otano A, del Río J (1998). Effect of chronic antidepressant treatment on transcription factor binding activity in rat hippocampus and frontal cortex. *Prog Neuropsychopharm Biol Psych* **22**: 787–802.

Frodl T, Meisensahl E, Zetzsche T, Bottlender R, Born C, Groll C *et al* (2002). Enlargement of the amygdala in patients with first episode of major depression. *Biol Psych* **51**: 708–714.

Fumagalli F, Frasca A, Sparta M, Drago F, Racagni G, Riva MA (2006). Long-term exposure to the atypical antipsychotic olanzapine differently up-regulates extracellular signal-regulated kinases 1 and 2 phosphorylation in subcellular compartments of rat prefrontal cortex. *Mol Pharm* **69**: 1366–1372.

Fumagalli F, Molteni R, Calabrese F, Frasca A, Racagni G, Riva MA (2005). Chronic fluoxetine administration inhibits extracellular signal-related kinase 1/2 phosphorylation in rat brain. *J Neurochem* **93**: 1551–1560.

Gallassi R, Di Sarro R, Morreale A, Amore M (2006). Memory impairment in patients with late-onset major depression: the effect of antidepressant therapy. *J Affect Disord* **91**: 243–250.

Giovannini MG (2006). The role of the extracellular signal-regulated kinase pathway in memory encoding. *Rev Neurosci* **17**: 619–634.

Glazewski S, Barth AL, Wallace H, McKenna M, Silva A, Fox K (1999). Impaired experience-dependent plasticity in barrel cortex of mice lacking the alpha and delta isoforms of CREB. *Cereb Cortex* **9**: 249–256.

Goldman-Rakic PS (1996). Regional and cellular fractionation of working memory. *Proc Natl Acad Sci USA* **93**: 13473–13480.

Gould E, Cameron HA, Daniels DC, Wooley CS, McEwen BS (1992). Adrenal hormones suppress cell division in the adult rat dentate gyrus. *J Neurosci* **12**: 3642–3650.

Gould E, Cameron HA (1997). Early NMDA receptor blockade impairs defensive behavior and increases cell proliferation in the dentate gyrus of developing rats. *Behav Neurosci* **11**: 49–56.

Grønli J, Bramham C, Murison R, Kanhema T, Fiske E, Bjorvatn B *et al* (2006). Chronic mild stress inhibits BDNF protein expression and CREB activation in the dentate gyrus but not in the hippocampus proper. *Pharmacol Biochem Behav* **85**: 842–849.

Hajszan T, MacLusky NJ, Leranth C (2005). Short-term treatment with the antidepressant fluxoetine triggers pyramidal dendritic spine synapse formation in rat hippocampus. *Eur J Neurosci* **21**: 1299–1303.

Hardingham GE, Bading H (2003). The yin and yang of NMDA receptor signaling. *Trends Neurosci* **26**: 81–89.

Hardingham GE, Fukunaga Y, Bading H (2002). Extrasynaptic NMDARs oppose synaptic NMDRs by triggering CREB shut-off and cell death pathways. *Nat Neurosci* **5**: 405–414.

Harvey PO, Fossati P, Pochon JB, Levy R, LeBastard G, Lehéricy S *et al* (2005). Cognitive control and brain resources in major depression: an fMRI study using the *n*-back task. *Neuroimage* **26**: 860–869.

Hasler G, Drevets WC, Manji HK, Charney DS (2004). Discovering endophenotypes for major depression. *Neuropsychopharmacology* **29**: 1765–1781.

Hebb DO (1949). *The Organization of Behavior: A Neuropsychological Theory*. Wiley: New York.

Heine VM, Zareno J, Maslam S, Joels M, Lucassen PJ (2005). Chronic stress in the adult dentate gyrus reduces cell proliferation near the vasculature and VEGF and Flk-1 protein expression. *Eur J Neurosci* **21**: 1304–1314.

Heldt SA, Stanek L, Chhatwal JP, Ressler KJ (2007). Hippocampus-specific deletion of BDNF in adult mice impairs spatial memory and extinction of aversive memories. *Mol Psych* **12**: 656–670.

Heninger GR, Delgado PL, Charney DS (1996). The revised monoamine theory of depression: a modulatory role for monoamines, based on new findings from monoamine depletion experiments in humans. *Pharmacopsychiatry* **29**: 2–11.

Holderbach R, Clark K, Moreau JL, Bischofberger J, Normann C (2007). Enhanced long-term synaptic depression in an animal model of depression. *Biol Psych* **62**: 92–100.

Hoshaw BA, Malberg JE, Lucki I (2005). Central administration of IGF-I and BDNF leads to long-lasting antidepressant effects. *Brain Res* **1037**: 204–208.

Huang YY, Martin KC, Kandel ER (2000). Both protein kinase A and mitogen-activated protein kinase are required in the amygdala for the macromolecular synthesis-dependent late phase of long-term potentiation. *J Neurosci* **20**: 6317–6325.

Jaffard R, Mocaer E, Poignant JC, Micheau J, Marighetto A, Meunier M *et al* (1991). Effects of tianeptine on spontaneous alternation, simple and concurrent spatial discrimination learning and on alcohol-induced alternation deficits in mice. *Behav Pharm* **2**: 37–46.

Kandel ER (2001). The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**: 1030–1038.

Kang H, Sun LD, Atkins CM, Soderling TR, Wilson MA, Tonegawa S (2001). An important role of neural activity-dependent CaMKIV signaling in the consolidation of long-term memory. *Neuron* **106**: 771–783.

Kaplan MS, Hinds JW (1977). Neurogenesis in the adult rat: electron microscopic analysis of light radioautographs. *Science* **197**: 1092–1094.

Karege F, Perret G, Bondolfi G, Schwald M, Bertschy G, Aubry JM (2002). Decreased serum brain-derived neurotrophic factor levels in major depressed patients. *Psych Res* **109**: 143–148.

Karege F, Vaudan G, Schwald M, Perroud N, La Harpe R (2005). Neurotrophin levels in postmortem brains of suicide victims and the effects of antemortem diagnosis and psychotropic drugs. *Brain Res Mol Brain Res* **136**: 29–37.

Karst H, Joëls M (2005). Corticosterone slowly enhances miniature excitatory postsynaptic current amplitude in mice CA1 hippocampal cells. *J Neurophys* **94**: 3479–3485.

Kasselman LJ, Kintner J, Sideris A, Pasnikowski E, Krellman JW, Shah S *et al* (2007). Dexamethasome treatment and ICAM-1 deficiency impair VEGF-induced angiogenesis in adult brain. *J Vasc Res* **44**: 283–291.

Kaufman J, Yang BZ, Douglas-Palumberi H, Grasso D, Lipschtz D, Houshyar S *et al* (2006). Brain-derived neurotrophic factor-5-HTTLPR gene interactions and environmental modifiers of depression in children. *Biol Psych* **59**: 673–680.

Kempermann G, Kuhn HG, Gage FH (1997). More hippocampal neurons in adult mice living in an enriched environment. *Nature* **386**: 493–495.

Kendler KS, Karkowski LM, Prescott CA (1999). Causal relationship between stressful life events and the onset of major depression. *Am J Psych* **156**: 837–841.

Kida S, Josselyn SA, de Ortiz SP, Kogan JH, Chevere I, Masushige S *et al* (2002). CREB required for the stability of new and reactivated fear memories. *Nat Neurosci* **5**: 348–355.

Kim JJ, Diamond DM (2002). The stressed hippocampus, synaptic plasticity and lost memories. *Nature Rev Neurosci* **3**: 453–462.

Kim JJ, Foy MR, Thompson RF (1996). Behavioral stress modifies hippocampal plasticity through *N*-methyl-aspartate receptor activation. *Proc Natl Acad Sci USA* **93**: 4750–4753.

Koch JM, Kell S, Hinze-Selch D, Aldenhoff JB (2002). Changes in CREB-phosphorylation during recovery from major depression. *J Psych Res* **36**: 369–375.

Lange C, Irle E (2004). Enlarged amygdala volume and reduced hippocampal volume in young women with major depression. *Psychol Med* **34**: 1059–1064.

Lee SY, Kang JS, Song GY, Myung CS (2006). Stress induces the expression of heterotrimeric G protein beta subunits and the phosphorylation of PKB/Akt and ERK1/2 in rat brain. *Neurosci Res* **56**: 180–192.

Leuner B, Gould E, Shors TJ (2006). Is there a link between adult neurogenesis and learning? *Hippocampus* **16**: 216–224.

Levkovitz Y, Grisaru N, Segal N (2001). Transcranial magnetic stimulation and antidepressive drugs share similar cellular effects in rat hippocampus. *Neuropsychopharmacology* **24**: 608–616.

Luine V, Martinez C, Villegas M, Magarinos A, McEwen B (1996). Restraint stress reversibly enhances spatial memory performance. *Physiol Behav* **59**: 27–32.

Lisman JE, Goldring MA (1988). Feasibility of long-term storage of graded information by the Ca^{2+}/calmodulin-dependent protein kinase molecules of the postsynaptic density. *Proc Natl Acad Sci USA* **85**: 5320–5324.

Lisman JE, Grace AA (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron* **46**: 703–713.

Liston C, Miller M, Goldwater DS, Radley JJ, Rocher AB, Hof PR *et al* (2006). Stress-induced alterations in prefrontal cortical dendritic morphology predict selective impairments in perceptual attentional set-shifting. *J Neurosci* **26**: 7870–7874.

Lowy MT, Gault L, Yamamoto BK (1993). Adrenalectomy attenuates stress-induced elevations in extracellular glutamate concentrations in the hippocampus. *J Neurochem* **61**: 1957–1960.

MacQueen GM, Campbell S, McEwen BS, Macdonald K, Amano S, Joffe RT *et al* (2003). Course of illness, hippocampal function, and hippocampal volume in major depression. *Proc Natl Acad Sci USA* **100**: 1387–1392.

Magariños AM, Deslandes A, McEwen BS (1999). Effects of antidepressants and benzodiazepine treatments on the dendritic structure of CA3 pyramidal neurons after chronic stress. *Eur J Pharmacol* **371**: 113–122.

Magarinos AM, McEwen BS, Flugge G, Fuchs E (1996). Chronic psychosocial stress causes apical dendritic atrophy of hippocampal CA3 pyramidal neurons in subordinate tree shrews. *J Neurosci* **16**: 3534–3540.

Magariños AM, McEwen BS (1995). Stress-induced atrophy of apical dendrites of hippocampal CA3c neurons: involvement of glucocorticoid secretion and excitatory amino acid receptors. *Neuroscience* **69**: 89–98.

Malberg JE, Eisch AJ, Nestler EJ, Duman RS (2000). Chronic antidepressant treatment increases neurogenesis in adult rat hippocampus. *J Neurosci* **20**: 9104–9110.

Malenka RC, Bear MF (2004). LTP and LTD: an embarrassment of riches. *Neuron* **44**: 5–21.

Malenka RC, Nicoll RA (1997). Silent synapses speak up. *Neuron* **19**: 473–476.

Malleret G, Haditsch U, Genoux D, Jones MW, Bliss TV, Vanhoose AM *et al* (2001). Inducible and reversible enhancement of learning, memory, and long-term potentiation by genetic inhibition of calcineurin. *Cell* **104**: 675–686.

Manji HK, Quiroz JA, Sporn J, Payne JL, Denicoff K, Gray NA *et al* (2003). Enhancing neuronal plasticity and cellular resilience to develop novel, improved therapeutics for difficult-to-treat depression. *Biol Psych* **53**: 707–742.

Maroun M, Richter-Levin G (2003). Exposure to acute stress blocks the induction of long-term potentiation of the amygdala-prefrontal cortex pathway *in vivo*. *J Neurosci* **23**: 4406–4409.

Maroun M (2006). Stress reverses plasticity in the pathway projecting from the ventromedial prefrontal cortex to the basolateral amygdala. *Eur J Neurosci* **24**: 2917–2922.

Massicotte G, Bernard J, Ohayon M (1993). Chronic effects of trimipramine, an antidepressant, on hippocampal synaptic plasticity. *Behav Neural Biol* **59**: 100–106.

Matsuzaki M, Honkura N, Ellis-Davies GC, Kasai H (2004). Structural basis of long-term potentiation in single dendritic spines. *Nature* **429**: 761–766.

McEwen BS (1999). Stress and hippocampal plasticity. *Ann Rev Neurosci* **22**: 105–122.

Metsis M, Timmusk T, Arenas E, Persson H (1993). Differential usage of multiple brain-derived neurotrophic factor promoters in the rat brain following neuronal activation. *Proc Natl Acad Sci USA* **90**: 8802–8806.

Miller RG, Mitchell JD, Lyon M, Moore DH (2007). Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *Cochrane Database Syst Rev* **1**: CD00147 (Review).

Ming G, Song H (2005). Adult neurogenesis in the mammalian central nervous system. *Ann Rev Neurosci* **28**: 223–250.

Moghaddam B, Adams B, Verma A, Daly D (1997). Activation of glutamatergic neurotransmission by ketamine: a novel step in the pathway from NMDA receptor blockade to dopaminergic and cognitive disruptions associated with the prefrontal cortex. *J Neurosci* **17**: 2921–2927.

Monteggia LM, Barrot M, Powell CM, Berton O, Galanis V, Gemelli T *et al* (2004). Essential role of brain-derived neurotrophic factor in adult hippocampal function. *Proc Natl Acad Sci USA* **101**: 10827–10832.

Monteggia LM, Luikart B, Barrot M, Theobold D, Malkovska I, Nef S *et al* (2007). Brain-derived neurotrophic factor conditional knockouts show gender differences in depression-related behaviors. *Biol Psych* **61**: 187–197.

Mowla A, Mosavinasab M, Pani A (2007). Does fluoxetine have any effects on the cognition of patients with mild cognitive

impairment? A double-blind, placebo-controlled, clinical trial. *J Clin Psychopharm* **27**: 67–70.

Muller D, Wang C, Skibo G, Toni N, Cremer H, Calaora V *et al* (1996). PSA-NCAM is required for activity-induced synaptic plasticity. *Neuron* **17**: 413–422.

Nair A, Vadodaria KC, Banerjee SB, Benekareddy M, Dias BG, Duman RS *et al* (2007). Stressor-specific regulation of distinct brain-derived neurotrophic factor transcripts and cyclic AMP response element-binding protein expression in the postnatal and adult rat hippocampus. *Neuropsychopharmacology* **32**: 1504–1519.

Nakagawa S, Kim JE, Lee R, Malberg JE, Chen J, Steffen C *et al* (2002a). Regulation of neurogenesis in adult mouse hippocampus by cAMP and the cAMP response element-binding protein. *J Neurosci* **22**: 3673–3682.

Nakagawa S, Kim JE, Lee R, Chen J, Fujioka T, Malberg J *et al* (2002b). Localization of phosphorylated cAMP response element-binding protein in immature neurons of adult hippocampus. *J Neurosci* **22**: 9868–9876.

Naudon L, Hotte M, Jay TM (2007). Effects of acute and chronic antidepressant treatment on memory performance: a comparison between paroxetine and imipramine. *Psychopharmacology (Berl)* **191**: 353–364.

Nemeroff CB (2006). Use of atypical antipsychotics in refractory depression and anxiety. *J Clin Psych* **66**(S8): 13–21.

Nestler EJ, Carlezon Jr WA (2006). The mesolimbic dopamine reward circuit in depression. *Biol Psych* **59**: 1151–1159.

Nestler EJ, Terwilliger RZ, Duman RS (1989). Chronic antidepressant administration alters the subcellular distribution of cyclic AMP-dependent protein kinase in rat frontal cortex. *J Neurochem* **53**: 1655–1657.

Newcomer J, Selke G, Melson A, Hershey T, Craft S, Richards K *et al* (1999). Decreased memory performance in healthy humans induced by stress-level cortisol treatment. *Arch Gen Psych* **56**: 527–533.

Newton SS, Colier EF, Hunsberger J, Adams D, Terwilliger R, Selvanayagam E *et al* (2003). Gene profile of electroconvulsive seizures: induction of neurotrophic and angiogenic factors. *J Neurosci* **23**: 10841–10851.

Newton SS, Duman RS (2004). Regulation of neurogenesis and angiogenesis in depression. *Curr Neurovasc Res* **1**: 261–267.

Newton SS, Thome J, Wallace TL, Shirayama Y, Schlesinger L, Sakai N *et al* (2002). Inhibition of cAMP response element-binding protein or dynorphin in the nucleus accumbens produces an antidepressant-like effect. *J Neurosci* **22**: 10883–10890.

Nguyen PV, Abel T, Kandel ER (1994). Requirement of a critical period of transcription for induction of a late phase of LTP. *Science* **265**: 1104–1107.

Nibuya M, Morinobu S, Duman RS (1995). Regulation of BDNF and TrkB mRNA in rat brain by chronic electroconvulsive seizure and antidepressant drug treatments. *J Neuosci* **15**: 7539–7547.

Nibuya M, Nestler EJ, Duman RS (1996). Chronic antidepressant administration increases the expression of cAMP response element binding protein (CREB) in rat hippocampus. *J Neurosci* **16**: 2365–2372.

Nibuya M, Takahashi M, Russell DS, Duman RS (1999). Repeated stress increases catalytic TrkB mRNA in rat hippocampus. *Neurosci Lett* **267**: 81–84.

Norrholm SD, Ouimet CC (2001). Altered dendritic spine density in animal models of depression and in response to antidepressant treatment. *Synapse* **42**: 151–163.

Nowakowska E, Kus K, Chodera A, Rybakowski J (2000). Behavioural effects of fluoxetine and tianeptine, two antidepressants with opposite action mechanisms, in rats. *Arzneimittelforschung* **50**: 5–10.

Nowakowska E, Kus K, Chodera A (2003). Comparison of behavioural effects of venlafaxine and imipramine in rats. *Arzneimittelforschung* **53**: 237–242.

Nowakowska E, Kus K, Florek E, Czubak A, Jodynis-Liebert J (2006). The influence of tobacco smoke and nicotine on antidepressant and memory-improving effects of venlafaxine. *Hum Exp Toxicol* **25**: 199–209.

O'Connor JJ, Rowan MJ, Anwyl R (1993). Naunyn Schmiedebergs Arch. *Pharmacology* **348**: 158–163.

Odagaki Y, García-Sevilla JA, Huguelet P, La Harpe R, Koyama T, Gulmón J (2001). Cyclic AMP-mediated signaling components are upregulated in the prefrontal cortex of depressed suicide victims. *Brain Res* **898**: 224–231.

O'Donnell JM, Zhang HT (2004). Antidepressant effects of inhibitors of cAMP phosphodiesterase (PDE4). *Trends Pharm Sci* **25**: 158–163.

Ongur D, Drevets WC, Price JL (1998). Glial reduction in the subgenual prefrontal cortex in mood disorders. *Proc Natl Acad Sci USA* **95**: 13290–13295.

Ortiz J, Fitzgerald LW, Lane S, Terwilliger R, Nestler EJ (1996). Biochemical effects of repeated stress in the mesolimbic dopamine system. *Neuropsychopharmacology* **14**: 393–402.

Pang PT, Lu B (2004). Regulation of late-phase LTP and long-term memory in normal and aging hippocampus: role of secreted proteins tPA and BDNF. *Ageing Res Rev* **3**: 407–430.

Pardon MC, Roberts ME, Marsden CA, Bianchi M, Latif ML, Duxon MS *et al* (2005). Social threat and novel cage stress-induced sustained extracellular-regulated kinase 1/2 (ERK 1/2) phosphorylation but differential modulation of brain-derived neurotrophic factor (BDNF) expression in the hippocampus of NMRI mice. *Neuroscience* **132**: 561–574.

Patterson SL, Abel T, Deuel TA, Martin KC, Rose JC, Kandel ER (1996). Recombinant BDNF rescues deficits in basal synaptic transmission and hippocampal LTP in BDNF knockout mice. *Neuron* **16**: 1137–1145.

Patterson SL, Grover LM, Schwartzkroin PA, Bothwell M (1992). Neurotrophin expression in rat hippocampal slices: a stimulus paradigm inducing LTP in CA evokes increases in BDNF and NT-3 mRNAs. *Neuron* **9**: 1081–1088.

Patterson SL, Pittenger C, Morozov A, Martin KC, Scanlin H, Drake C *et al* (2001). Some forms of cAMP-mediated long-lasting potentiation are associated with release of BDNF and nuclear translocation of phospho-MAP kinase. *Neuron* **32**: 123–140.

Paul IA, Skolnick P (2003). Glutamate and depression: clinical and preclinical studies. *Ann NY Acad Sci* **1003**: 250–272.

Perez J, Tardito D, Mori S, Racagni G, Smeraldi E, Zanardi R (2000). Abnormalities of cAMP signaling in affective disorders: implication for pathophysiology and treatment. *Bipolar Disord* **2**: 27–36.

Perez J, Tinelli D, Brunello N, Racagni G (1989). cAMP-dependent phosphorylation of soluble and crude microtubule fractions of rat cerebral cortex after prolonged desmethylimipramine treatment. *Eur J Pharmacol* **172**: 305–316.

Perlis RH, Purcell S, Fagerness J, Cusin C, Yamaki L, Fava M *et al* (2007). Clinical and genetic dissection of anger expression and CREB1 polymorphisms in major depressive disorder. *Biol Psych* **62**: 536–540.

Pickering C, Gustafsson L, Cebere A, Nylander I, Liljequist S (2006). Repeated maternal separation of male Wistar rats alters glutamate receptor expression in the hippocampus but not the prefrontal cortex. *Brain Res* **1099**: 101–108.

Pittenger C, Fasano S, Mazzocchi-Jones D, Dunnett SB, Kandel ER, Brambilla R (2006). Impaired bidirectional synaptic plasticity and procedural memory formation in striatum-specific cAMP response element-binding protein-deficient mice. *J Neurosci* **26**: 2808–2813.

Pittenger C, Huang YY, Paletzki RF, Bourtchouladze R, Scanlin H, Vronskaya S; *et al* (2002). Reversible inhibition of CREB/ATF transcription factors in region CA1 of the dorsal hippocampus disrupts hippocampus-dependent spatial memory. *Neuron* **34**: 447–462.

Pittenger C, Sanacora G, Krystal JH (2007). The NMDA receptor as a therapeutic target in major depressive disorder. *CNS Neurol Disord—Drug Targets* **6**: 101–115.

Pliakas AM, Carlson RR, Neve RL, Konradi C, Nestler EJ, Carlezon Jr WA (2001). Altered responsiveness to cocaine and increased immobility in the forced swim test associated with elevated cAMP response element-binding protein expression in nucleus accumbens. *J Neurosci* **21**: 7397–7403.

Popoli M, Vocaturo C, Perez J, Smeraldi E, Racagni G (1995). Presynaptic Ca^{2+}/calmodulin-dependent protein kinase II: autophosphorylation and activity increase in the hippocampus after long-term blockade of serotonin reuptake. *Mol Pharm* **48**: 623–629.

Radley JJ, Sisti HM, Hao J, Rocher AB, McCall T, Hof PR *et al* (2004). Chronic behavioral stress induces apical dendrite reorganization in pyramidal neurons of the medial prefrontal cortex. *Neuroscience* **125**: 1–6.

Rajkowska G, Miguel-Hidalgo JJ, Wei J, Dilley G, Pittman SD, Meltzer HY *et al* (1999). Morphometric evidence for neuronal and glial prefrontal cell pathology in major depression. *Biol Psych* **45**: 1085–1098.

Rajkowska G, O'Dwyer G, Teleki Z, Stockmeier CA, Miguel-Hidalgo JJ (2007). GABAergic neurons immunoreactive for calcium binding proteins are reduced in the prefrontal cortex in major depression. *Neuropsychopharmacology* **32**: 471–482.

Rasmusson AM, Shi L, Duman R (2002). Downregulation of BDNF mRNA in the hippocampal dentate gyrus after re-exposure to cues previously associated with footshock. *Neuropsychopharmacology* **27**: 133–142.

Rattiner LM, Davis M, Ressler KJ (2005). Brain-derived neurotrophic factor in amygdala-dependent learning. *Neuroscientist* **11**: 323–333.

Reif A, Fritzen S, Finger M, Strobel A, Lauer M, Schmitt A *et al* (2006). Neural stem cell proliferation is decreased in schizophrenia, but not in depression. *Mol Psych* **11**: 514–522.

Revest JM, Di Blasi F, Kitchener P, Rougé-Pont F, Desmedt A, Turiault M *et al* (2005). The MAPK pathway and Egr-1 mediate stress-related behavioral effects of glucocorticoids. *Nat Neurosci* **8**: 664–672.

Reynolds IJ, Miller RJ (1988). Tricyclic antidepressants block *N*-methyl-D-aspartate receptors: similarities to the action of zinc. *Br J Pharmacol* **95**: 95–102.

Rocher C, Spedding M, Munoz C, Jay TM (2004). Acute stress-induced changes in hippocampal/prefrontal circuits in rats: effects of antidepressants. *Cereb Cortex* **14**: 224–229.

Russo-Neustadt A, Ha T, Ramirez R, Kesslak JP (2001). Physical activity-antidepressant treatment combination: impact on brain-derived neurotrophic factor and behavior in an animal model. *Behav Brain Res* **120**: 87–95.

Saal D, Dong Y, Bonci A, Malenka RC (2003). Drugs of abuse and stress trigger a common synaptic adaptation in dopamine neurons. *Neuron* **37**: 577–582.

Saarelainen T, Hendolin P, Lucas G, Koponen E, Sairanen M, MacDonald E *et al* (2003). Activation of the TrkB neurotrophin receptor is induced by antidepressant drugs and is required for antidepressant-induced behavioral effects. *J Neurosci* **23**: 349–357.

Sackeim HA, Freeman J, McElhiney M, Coleman E, Prudic J, Devanand DP (1992). Effects of major depression on estimates of intelligence. *J Clin Exp Neuropsychol* **14**: 268–288.

Sackeim HA (2000). Memory and ECT: from polarization to reconciliation. *J ECT* **16**: 87–96.

Sairanen M, Lucas G, Ernfors P, Castren M, Casren E (2005). Brain-derived neurotrophic factor and antidepressant drugs have different but coordinated effects on neuronal turnover, proliferation, and survival in the adult dentate gyrus. *J Neurosci* **25**: 1089–1094.

Salm AK, Pavelko M, Krouse EM, Webster W, Kraszpulski M, Birkle DL (2004). Lateral amygdaloid nucleus expansion in adult rats is associated with exposure to prenatal stress. *Brain Res Dev Brain Res* **148**: 159–167.

Sanacora G, Kendell SF, Levin Y, Simen AA, Fenton LR, Coric V *et al* (2007). Preliminary evidence of riluzole efficacy in antidepressant-treated patients with residual depressive symptoms. *Biol Psych* **61**: 822–825.

Sanes JR, Lichtman JW (1999). Can molecules explain long-term potentiation? *Nat Neurosci* **2**: 597–604.

Santarelli L, Saxe M, Gross C, Surget A, Battaglia F, Dulawa S *et al* (2003). Requirement of hippocampal neurogenesis for the behavioral effects of antidepressants. *Science* **301**: 805–809.

Sapolsky RM (2003). Stress and plasticity in the limbic system. *Neurochem Res* **28**: 1735–1742.

Sapolsky RM, Krey L, McEwen B (1985). Prolonged glucocorticoid exposure reduces hippocampal neuron number: implications for aging. *J Neurosci* **5**: 1221–1227.

Sapolsky RM (2000). Glucocorticoids and hippocampal atrophy in neuropsychiatric disorders. *Arch Gen Psych* **57**: 925–935.

Saxe MD, Battaglia F, Wang JW, Malleret G, David DJ, Monckton JE *et al* (2006). Ablation of hippocampal neurogenesis impairs contextual fear conditioning and synaptic plasticity in the dentate gyrus. *Proc Natl Acad Sci USA* **103**: 17501–17506.

Schaaf MJ, de Kloet ER, Vreugdenhil E (2000). Corticosterone effects on BDNF expression in the hippocampus. Implications for memory formation. *Stress* **3**: 201–208.

Schlaepfer TE, Cohen MX, Frick C, Kosel M, Brodesser D, Axmacher N *et al* (2007). Deep brain stimulation to reward circuitry alleviates anhedonia in refractory major depression. *Neuropsychopharmacology*; epub ahead of print.

Shah PJ, Ebmeier KP, Glabus MF, Goodwin GM (1998). Cortical grey matter reductions associated with treatment-resistant chronic unipolar depression. Controlled magnetic resonance imaging study. *Br J Psych* **172**: 527–532.

Sheline YI, Sanghavi M, Mintun MA, Gado MH (1999). Depression duration but not age predicts hippocampal volume loss in medically healthy women with recurrent major depression. *J Neurosci* **19**: 5034–5043.

Shimizu E, Hashimoto K, Okamura N, Koike K, Komatsu N, Kumakiri C *et al* (2003). Alterations of serum levels of brain-derived neurotrophic factor (BDNF) in depressed patients with or without antidepressants. *Biol Psych* **54**: 70–75.

Shirayama Y, Chen AC, Nakagaway S, Russell DS, Duman RS (2002). Brain-derived neurotrophic factor produces antidepressant effects in behavioral models of depression. *J Neurosci* **22**: 3251–3261.

Shors TJ (2001). Acute stress rapidly and persistently enhances memory formation in the male rat. *Neurobiol Learn Mem* **75**: 10–29.

Shors TJ (2005). Stressful experience and learning across the lifespan. *Ann Rev Psychol* **57**: 55–85.

Shors TJ, Miesegaes G, Beylin A, Zhao M, Rydel T, Gould E (2001). Neurogenesis in the adult is involved in the formation of trace memories. *Nature* **410**: 372–376.

Shors TJ, Seib TB, Levine S, Thompson RF (1989). Inescapable versus escapable shock modulates long-term potentiation in the rat hippocampus. *Science* **244**: 224–226.

Silva-Gomez AB, Rojas D, Juarez I, Flores G (2003). Decreased dendritic spine density on prefrontal cortical and hippocampal pyramidal neurons in postweaning social isolation rats. *Brain Res* **983**: 128–136.

Siuciak JA, Lewis DR, Wiegand SJ, Lindsay RM (1997). Antidepressant-like effect of brain-derived neurotrophic factor (BDNF). *Pharmacol Biochem Behav* **56**: 131–137.

Smith MA, Makino S, Kvetnansky R, Post RM (1995). Stress and glucocorticoids affect the expression of brain-derived neurotrophic factor and neurotrophin-3 mRNAs in the hippocampus. *J Neurosci* **15**: 1768–1777.

Snyder JS, Hong NS, McDonald RJ, Wojtowicz JM (2005). A role for adult neurogenesis in spatial long-term memory. *Neuroscience* **130**: 843–852.

Squire LR, Stark CE, Clark RE (2004). The medial temporal lobe. *Ann Rev Neurosci* **27**: 279–306.

Stewart CA, Reid IC (2000). Repeated ECS and fluoxetine administration have equivalent effects on hippocampal synaptic plasticity. *Psychopharmacology (Berl)* **148**: 217–223.

Stockmeier CA, Mahajan GJ, Konick LC, Overholser JC, Jurjus GJ, Meltzer HY *et al* (2004). Cellular changes in the postmortem hippocampus in major depression. *Biol Psych* **56**: 640–650.

Svenningsson P, Tzavara ET, Witkin JM, Fienberg AA, Nomikos GG, Greengard P (2002). Involvement of striatal and extrastriatal DARP-32 in biochemical effects of fluoxetine (Prozac). *Proc Natl Acad Sci USA* **99**: 3182–3187.

Tardito D, Perez J, Tiraboschi E, Musazzi L, Racagni G, Popoli M (2006). Signaling pathways regulating gene expression, neuroplasticity, and neurotrophic mechanisms in the action of antidepressants: a critical overview. *Pharmacol Rev* **58**: 115–134.

Tashiro A, Makino H, Gage FH (2007). Experience-specific functional modification of the dentate gyrus through adult neurogenesis: a critical period during an immature stage. *J Neurosci* **27**: 3252–3259.

Thome J, Sakai N, Shin K, Steffen C, Zhang YJ, Impey S *et al* (2000). cAMP response element-mediated gene transcription is upregulated by chronic antidepressant treatment. *J Neurosci* **20**: 4030–4036.

Tiraboschi E, Giambelli R, D'Urso G, Galietta A, Barbon A, de Bartolomeis A *et al* (2004a). Antidepressants activate CaMKII in neuron cell body by Thr286 phosphorylation. *Neuroreport* **15**: 2393–2396.

Tiraboschi E, Tardito D, Kasahara J, Moraschi S, Pruneri P, Genarelli M *et al* (2004b). Selective phosphorylation of nuclear CREB by fluoxetine is linked to activation of CaMK IV and MAP kinase cascades. *Neuropsychopharm* **29**: 1831–1840.

Toni N, Teng EM, Bushong EA, Aimone JB, Zhao C, Consiglio A *et al* (2007). Synapse formation on neurons born in the adult hippocampus. *Nat Neurosci* **10**: 727–734.

Tononi G, Cirelli C (2006). Sleep function and synaptic homeostasis. *Sleep Med Rev* **10**: 49–62.

Uranova NA, Vostrikov VM, Orlovskaya DD, Rachmanova VI (2004). Oligodendroglial density in the prefrontal cortex in schizophrenia and mood disorders: a study from the Stanley Neuropathology Consortium. *Schizophr Res* **67**: 269–275.

Van Praag H, Kempermann G, Gage FH (2000). Neural consequences of environmental enrichment. *Nat Rev Neurosci* **1**: 191–198.

Venero C, Borrell J (1999). Rapid glucocorticoid effects on excitatory amino acid levels in the hippocampus: a microdialysis study in freely moving rats. *Eur J Neurosci* **11**: 2465–2473.

Vermetten E, Vythilingam M, Southwick SM, Charney DS, Bremner JD (2003). Long-term treatment with paroxetine increases verbal declarative memory and hippocampal volume in posttraumatic stress disorder. *Biol Psychiatry* **54**: 693–702.

Von Frijtag JC, Kamal A, Reijmers LG, Schrama LH, van den Bos R, Spruijt BM (2001). Chronic imipramine treatment partially reverses the long-term changes of hippocampal synaptic plasticity in socially stressed rats. *Neurosci Lett* **309**: 153–156.

Vouimba RM, Munoz C, Diamond DM (2006). Differential effects of predator stress and the antidepressant tianeptine on physiological plasticity in the hippocampus and basolateral amygdala. *Stress* **9**: 29–40.

Vyas A, Bermal S, Chattarji S (2003). Effects of chronic stress on dendritic arborization in the central and extended amygdala. *Brain Res* **965**: 290–294.

Vyas A, Jadhav S, Chattarji S (2006). Prolonged behavioral stress enhances synaptic connectivity in the basolateral amygdala. *Neuroscience* **143**: 387–393.

Vyas A, Mitra R, Shankaranarayana Rao BS, Chattarji S (2002). Chronic stress induces contrasting patterns of dendritic remodeling in hippocampal and amygdaloid neurons. *J Neurosci* **22**: 6810–6818.

Vyas A, Pillai AG, Chatarji S (2004). Recovery after chronic stress fails to reverse amygdaloid neuronal hypertrophy and enhanced anxiety-like behavior. *Neuroscience* **128**: 667–673.

Vythilingam M, Vermetten E, Anderson GM, Luckenbaugh D, Anderson ER, Snow J *et al* (2004). Hippocampal volume, memory, and cortisol status in major depressive disorder: effects of treatment. *Biol Psych* **56**: 101–112.

Warner-Schmidt JL, Duman RS (2006). Hippocampal neurogenesis: opposing effects of stress and antidepressant treatment. *Hippocampus* **16**: 239–249.

Warner-Schmidt JL, Duman RS (2007). VEGF is an essential mediator of the neurogenic and behavioral actions of antidepressants. *Proc Natl Acad Sci USA* **104**: 4647–4652.

Watanabe Y, Gould E, McEwen BS (1992). Stress induces atrophy of apical dendrites of hippocampal CA3 pyramidal neurons. *Brain Res* **588**: 341–345.

Wellman CL (2001). Dendritic reorganization in pyramidal neurons in medial prefrontal cortex after chronic corticosterone administration. *J Neurobiol* **49**: 245–253.

Willner P (2005). Chronic mild stress (CMS) revisited: consistency and behavioural-neurobiological concordance in the effects of CMS. *Neuropsychobiology* **52**: 90–110.

Witkin JM, Marek GJ, Johnson BG, Schoepp DD (2007). Metabotropic glutamate receptors in the control of mood disorders. *CNS Neurol Disord—Drug Targets* **6**: 87–100.

Woolley C, Gould E, McEwen B (1990). Exposure to excess glucocorticoids alters dendritic morphology of adult hippocampal pyramidal neurons. *Brain Res* **521**: 225–231.

Xu L, Anwyl R, Rowan MJ (1997). Behavioural stress facilitates the induction of long-term depression in the hippocampus. *Nature* **387**: 497–500.

Yamada S, Yamamoto M, Ozawa H, Riederer P, Saito T (2003). Reduced phosphorylation of cyclic AMP-responsive element binding protein in the postmortem oribitofrontal cortex of patients with major depressive disorder. *J Neural Transm* **110**: 671–680.

Yang CH, Huang CC, Hsu KS (2004). Behavioral stress modifies hippocampal plasticity through corticosterone-induced sustained extracellular signal-regulate kinase/mitogen-activated protein kinase activation. *J Neurosci* **24**: 11029–11034.

Yau JL, Noble J, Hibberd C, Rowe WB, Meaney MJ, Morris RG *et al* (2002). Chronic treatment with the antidepressant amitriptyline prevents impairments in water maze learning in aging rats. *J Neurosci* **22**: 1436–1442.

Zakzanis KK, Leach L, Kaplan E (1998). On the nature and pattern of neurocognitive function in major depressive disorder. *Neuropsych Neuropsychol Behav Neurol* **11**: 111–119.

Zarate Jr CA, Payne JL, Quiroz J, Sporn J, Denicoff KK, Luckenbaugh D *et al* (2004). An open-label trial of riluzole in patients with treatment-resistant major depression. *Am J Psych* **161**: 171–174.

Zarate CA, Singh JB, Carlson PJ, Brutsche NE, Ameli R, Luckenbaugh DA *et al* (2006). A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. *Arch Gen Psychiatry* **63**: 856–864.

Zschocke J, Bayatti N, Clement AM, Witan H, Figiel M, Engele J *et al* (2005). Differential promotion of glutamate transporter expression and function by glucocorticoids in astrocytes from various brain regions. *J Biol Chem* **280**: 34924–34932.

Zubenko GS, Hughes III HB, Stiffler JS, Brechbiel A, Zubenko WN, Maher BS *et al* (2003). Sequence variations in CREB1 cosegregate with depressive disorders in women. *Mol Psych* **8**: 611–618.

American College of Neuropsychopharmacology

www.nature.com/npp

ARTICLE

Stress-induced generalization of negative memories is mediated by an extended hippocampal circuit

Lynn Y. Ren[1]✉, Mariah A. A. Meyer[1], Viktoriya S. Grayson[1], Pan Gao[1], Anita L. Guedea[1] and Jelena Radulovic[1,2,3]✉

© The Author(s), under exclusive licence to American College of Neuropsychopharmacology 2021

Memories of negative experiences exert important control of behavior in the face of actual or anticipated threat. Sometimes, however, this control extends to non-threatening situations, a phenomenon known as overgeneralization of negative memories. Overgeneralization is a reliable cognitive phenotype of major depressive disorder, generalized anxiety disorder, and post-traumatic stress disorder. We therefore sought to develop an animal model to study stress-induced generalization of negative memories (SIG) and determine its dependence on the episodic-like memory circuit. We found that male and female mice, which were trained to differentiate a threatening from neutral context, exhibited robust SIG in response to subsequent social stress. Using chemogenetic circuit manipulations during memory retrieval, we demonstrated that both excitatory afferents to the dorsal hippocampus (DH) from the ventral tegmental area (VTA), and excitatory efferents from the DH to the retrosplenial cortex (RSC) contribute to SIG. Based on the known roles of these projections, we suggest that (1) by targeting subcortical VTA circuits that provide valence signals to the DH, stress prioritizes the retrieval of negative over neutral memories, and (2) by forwarding such information to the RSC, stress engages cortical mechanisms that support the retrieval of general relative to specific memory features. Altogether, these results suggest that various components of the extended hippocampal circuit can serve as treatment targets for memory overgeneralization.

Neuropsychopharmacology; https://doi.org/10.1038/s41386-021-01174-4

INTRODUCTION

Generalized anxiety disorder (GAD), major depressive disorder (MDD), and post-traumatic stress disorder (PTSD) are often accompanied by debilitating cognitive symptoms, which impede recovery from these disorders, often persist after antidepressant treatment of other symptoms, and increase the likelihood of relapse [1–7]. Additionally, they precede and predict associated affective symptoms [1, 8, 9]. A recurrent cognitive symptom across MDD, GAD, and PTSD is overgeneralization of negative memories [10–20], which involves excessive control of behavior by negative memories, even in neutral situations. Stress can promote generalization of negative memories [21, 22]; however, the neurobiological and circuit mechanisms underlying this effect are not well-understood. Notably, current knowledge of memory generalization focuses largely on processes related to the encoding [23–27] of new memories in a manner that promotes their generalization, whereas much less is known about memory retrieval processes that act upon consolidated memories [28, 29].

The dorsal hippocampus (DH) is crucial for mediating valence-memory integration, memory formation, memory retrieval, and memory generalization [30–38]. These functions partly rely on vGlut2-positive afferents from the ventral tegmental area (VTA) that provide a valence signal to the DH [33] and DH vGlut1-positive efferents to the retrosplenial cortex (RSC), which contribute to the long-lasting representations of episodic-like memories [39–45]. We therefore hypothesized that these components of the extended hippocampal circuit might contribute to stress-induced generalization (SIG).

Chronic social stressors, such as social defeat stress and social instability stress, induce a spectrum of anxiety-like and depressive-like behaviors in rodents [46–54]. Here, we demonstrate that exposure to subacute versions of these social stressors induces SIG in both male and female mice that is not associated with stress-induced anxiety-like and depressive-like behavioral phenotypes. Using chemogenetic manipulations, we show that SIG occurs at memory retrieval and is mediated by circuits involved in processing negative valence (VTA to DH) and long-term memory representation (DH to RSC).

METHODS

Mice

We performed these experiments using male and female C57BL/6J mice, vGlut1-Cre, and vGlut2-Cre mice. Wild-type C57BL/6J mice were purchased from Harlan (Indianapolis, IN) and wild-type CD-1 aggressor male mice were purchased from Charles Rivers (Wilmington, MA). All other mouse lines were obtained from Jackson Laboratory (Bar Harbor, ME). The vGlut1-Cre knockin mice, also known as Slc17a7-IRES2-Cre-D and Vglut1-IRES2-Cre-D, express Cre-recombinase in glutamatergic neuron cell bodies containing vesicular glutamate transporter 1 (vGlut1). The vGlut2-Cre knockin mice, also known as Slc17a6tm2(cre)Lowl and Vglut2-IRES-Cre, express Cre-recombinase in glutamatergic neuron cell bodies containing vesicular glutamate transporter 2 (vGlut2). Heterozygous mice were

[1]Department of Psychiatry and Behavioral Sciences, Northwestern University, Chicago, IL, USA. [2]Department of Psychiatry and Behavioral Sciences, Albert Einstein College of Medicine, Bronx, NY, USA. [3]Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA. ✉email: lynn.ren@northwestern.edu; jelena.radulovic@einsteinmed.org

Received: 28 April 2021 Revised: 20 August 2021 Accepted: 29 August 2021
Published online: 07 September 2021

backcrossed with wild-type C57BL/6J for six generations in our facility to achieve offspring with a genetic identity closer to the C57BL/6J strain. The colony was subsequently expanded by homozygous breeding and confirmed by genotyping using primers and protocols posted on the Jackson Laboratory website.

All mice were 8 weeks old at the beginning of the experiments. The mice were maintained under standard housing conditions (12/12 h light–dark cycle with lights on at 7 a.m., temperature 20–22 °C, humidity 30–60%) in our satellite behavioral facility. Randomization was performed by assigning similar numbers of littermates to the different treatment conditions. Behavioral experiments were performed by an experimenter blind to drug treatments and experimental groups. All animal procedures used in this study were approved by the Northwestern University IACUC and complied with federal regulations set forth by the National Institutes of Health.

Contextual fear conditioning and generalization

Contextual fear conditioning was performed in an automated system (TSE Systems, Chesterfield, MO) [45]. Context A (Cxt A) was scented with ethanol, had a metal grid floor, was rectangular in shape, and had dim room lighting. Context B (Cxt B) was scented with acetic acid, had a smooth plastic floor, was oval in shape, and had brighter room lighting. Cxts A and B were cleaned with 70% ethanol and 1% acetic acid, respectively, between mice. Mice were pre-exposed to Cxt A and Cxt B for 3 min each. Exposures to these contexts were spaced 3 h apart, with half of the mice in each experimental group being exposed to Cxt A first and the other half being exposed to Cxt B first. 24 h later, mice underwent 3 days of contextual fear conditioning. Each day consisted of 3 min of exposure to the conditioning context, followed by a footshock (2 s, 0.7 mA, constant current). Twenty-four hours after the last day of contextual fear conditioning, mice were tested for memory generalization by placing them into Cxt B for 3 min, as adapted from Xu et al. [55]. Freezing was scored every 10 s during context exposures and expressed as a percentage of the total number of observations during which the mice were motionless. Mice were also tested for memory retrieval by returning them to Cxt A for 3 min. The Generalization Index was adapted from discrimination indices that are commonly used to assess context discrimination [24, 55] and was calculated as follows:

$$0.5 - [(\text{Cxt A freezing} - \text{Cxt B freezing})/(\text{Cxt A freezing} + \text{Cxt B freezing})]$$

A higher generalization index indicated increased generalization.

Mice were exposed to social defeat stress or social instability stress (see below for details) 24 h after pre-stress memory tests. Twenty-four hours after the completion of social defeat stress or social instability stress, mice underwent post-stress memory tests.

Social defeat stress and social instability stress

Social defeat stress (SDS) [51] was conducted in male mice by placing an "intruder" (the experimental mouse) into the home cage of an aggressive "resident" (CD-1 male mouse) for 5 min. Intruder mice were returned to their own home cage after 5 min. Experimental mice underwent social defeat once per day for 4 consecutive days and were exposed to a different aggressive resident mouse each day.

Social Instability Stress (SIS) [52, 53] was conducted in female mice by alternating mice between 72 h of housing with three other unfamiliar females and 24 h of isolation. During each round of 72-h housing, the mice were exposed to three new cage-mates. This continued for a total of 11 days.

Elevated plus maze test, open-field test, sucrose preference test, and light–dark emergence test

During the sucrose preference test (SPT), mice were allowed to freely drink from two 50 mL graduated bottles. One bottle contained 1% sucrose (vol/vol) in distilled water, and the other bottle contained distilled water alone. Mice underwent 48 h of habituation to these bottles, during which bottle locations were alternated after 24 h. After habituation, sucrose and water consumption were measured every 24 h for a total of 96 h. The mass of liquid consumed was measured to a resolution of ±0.01 g. Water and sucrose bottle locations were alternated after every measurement to eliminate position bias.

During the open-field test (OFT), mice were placed in the center of a 50 × 50 cm platform for 5 min. The time spent in the center region (25 × 25 cm), velocity, and total distance traveled were measured using automatic video tracking software (VideoMot, TSE Systems). The platform was cleaned with 0.5% ammonium between mice.

During elevated plus maze (EPM) testing, mice were placed in the center of a plus shaped platform with two open arms and two closed arms for 5 min. The dimensions of each arm were 30 × 5 cm. The time spent in the open and closed arms, velocity, and total distance traveled were measured using automatic video tracking software (VideoMot). The platform was cleaned with 0.5% ammonium between mice.

During light–dark emergence (LDE), mice were placed into a dark box (10 ×10 x 10 cm) in the center of a 50 ×50 cm platform and the latency for the mice to emerge from the dark box was measured. The platform was cleaned with 0.5% ammonium between mice.

Stereotaxic surgeries and infusions of viral vectors and drugs

Mice were anesthetized with 1.2% tribromoethanol (vol/vol; Avertin) for viral vector intracranial infusion and cannula implantation. AAV8-hSyn-DIO-GFP (Addgene 50457, Addgene, Watertown, MA), AAV8-hSyn-DIO-mCherry (Addgene 50459), or AAV8-hSyn-DIO-hM4D(Gi)-mCherry (Addgene 50475) were bilaterally infused into the VTA (3.4 mm posterior, ±0.5 mm lateral, 4.3 mm ventral to bregma) or DH (1.8 mm posterior, ±1 mm lateral, 2.25 mm ventral to bregma). Infusions were performed using an automatic pump controller (Micro4-WPI, Sarasota, FL) connected to a microsyringe (Hamilton, Reno, NV). Viral vectors were infused into the VTA at a volume of 0.35 µL per site or into the DH at 0.5 µL per site, at a titer $\geq 3 \times 10^{12}$ vg/mL, over 2 min. Syringes were left in place for 5 min to allow for virus diffusion. Mice were allowed 4 weeks for virus expression before behavioral experiments.

Bilateral 26-gauge guide cannulas (Plastics One, Roanoke, VA) were placed in the DH (1.8 mm posterior, ±1.0 mm lateral, 2.25 mm ventral to bregma) or RSC (1.8 mm posterior, ±0.4 mm lateral, 0.75 mm ventral to bregma). Mice were allowed 1 week of recovery before behavioral experiments. Clozapine-N-Oxide (CNO; Sigma, St. Louis, MO; 0.3 µg/mL in 0.7% DMSO and 99.3% aCSF (vol/vol)) or vehicle (Veh; 0.7% DMSO and 99.3% aCSF (vol/vol)) was infused through the cannulas (0.25 µL per side into DH or 0.20 µL per side into RSC over 30 s) 30 min prior to context retrieval or the generalization test. After the completion of behavioral testing, mice were intracardially perfused and cannula placements and virus spread were confirmed by immunohistochemical analysis of mCherry signals.

Immunohistochemistry and immunofluorescence

Mice were anesthetized with an i.p. injection of 240 mg/kg Avertin and transcardially perfused with ice-cold 4% paraformaldehyde in phosphate buffer (pH 7.4). Brains were removed and post-fixed for 24 h in the same fixative solution and then immersed for 24 h each in 20% and 30% sucrose in phosphate buffer. Brains was frozen and 50 µm sections were cut for use in free-floating immunohistochemistry with primary antibodies against mCherry (1:16000, chicken polyclonal, ab205402, abcam, Cambridge, MA) or tyrosine hydroxylase (1:2000, mouse monoclonal, Immunostar 22941, Immunostar, Hudson, WI) and biotinylated goat anti chicken IgG or biotinylated goat anti mouse IgG secondary antibodies (1:200, Vector Laboratories, Burlingame, CA). Signals were amplified using ABC-HRP (1:100, Vector Laboratories) and visualized with TSA tetramethylrhodamine (1:62.5, Perkin Elmer, Waltham, MA) using fluorescent microscopy.

Quantification and statistical analyses

Statistical analyses were performed using Graphpad (San Diego, CA). All statistical tests were two-sided. Data used for analyses included: (1) mice with virus expression in the DH or VTA of 50% or more of the mean number of DH or VTA-infected neurons across all experiments, (2) mice with detectable VTA-originating terminals in the DH or DH-originating terminals in the RSC, and (3) mice with correctly implanted cannulas in the DH or RSC. For the behavioral studies, two-way RM ANOVA using either (1) context and treatment as factors or (2) context and stress as factors was used to determine differences in conditioning or altered context freezing between groups, as well as the effect of stress on conditioning context and altered context freezing. Significant *F*-values were followed by post-hoc multiple comparisons using Tukey's or Sidak's test. To calculate the generalization index, the same mice were exposed to both Cxt A and Cxt B. Differences in the generalization index were also determined for treatment (CNO or Veh) or stress (Pre-Stress vs Post-Stress) using two-tailed Mann–Whitney test or Wilcoxon matched-pairs signed rank test. Differences in behavioral measures during EPM, OFT, LDE, or SPT paradigms were determined for stress (No Stress vs Stress) using two-tailed Students T test. Statistical differences were considered significant for all *p*-values <

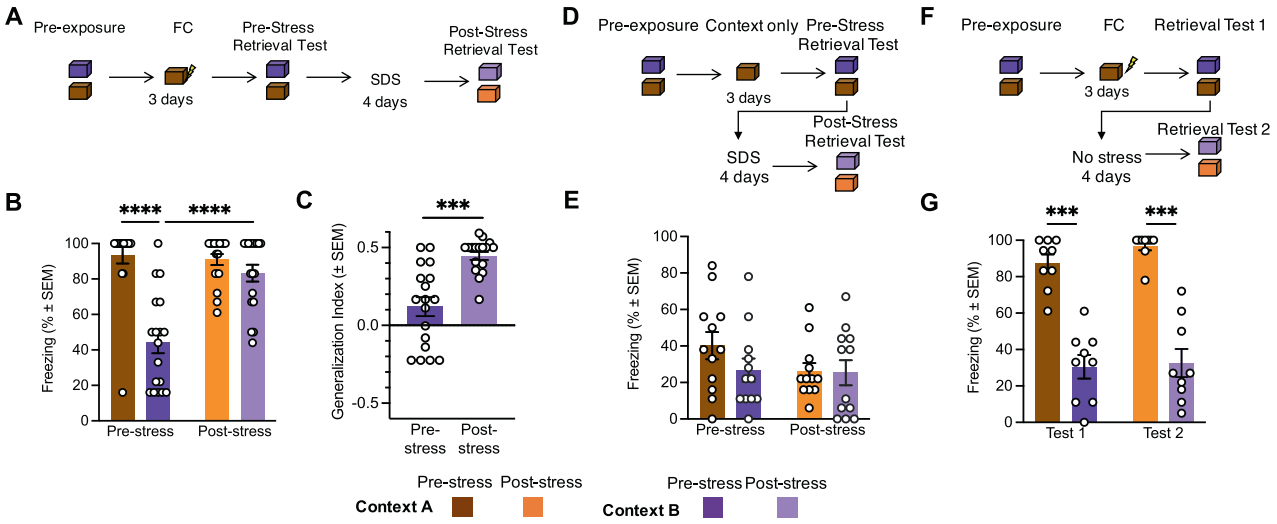


Fig. 1 Stress induces generalization of negative memories in male mice. A Behavioral schematic for SIG performed in male mice and shown in **B**, **C**. FC, fear conditioning. Arrows indicate the passage of 24 h. **B** Freezing in Context A and Context B in male mice before and after SDS. Two-way RM ANOVA. Factor: Stress, $F(1,17) = 15.15$, $p = 0.0012$. Factor: Context, $F(1,17) = 44.66$, $p < 0.0001$. Factor: Stress x Context, $F(1,17) = 34.83$, $p < 0.0001$. Post-hoc Tukey ****$p < 0.0001$. $n = 18$ mice per group. **C** Generalization index (see methods for formula) in male mice before and after SDS. Wilcoxon matched-pairs signed rank test, ***$p = 0.0001$. $n = 18$ mice per group. **D** Behavioral schematic for **E**. Note that mice were exposed to context without footshock. **E** Freezing in Context A and Context B after SDS in the absence of FC. Two-way RM ANOVA. Factor: Stress, $F(1,11) = 1.702$, $p = 0.2187$. Factor: Context, $F(1,11) = 1.032$, $p = 0.3315$. Factor: Stress x Context, $F(1,11) = 0.8979$, $p = 0.3637$. $n =$ 12 mice per group. **F** Behavioral schematic for **G**. Note that mice were exposed to FC but not SDS. **G** Freezing in Context A and Context B after FC in the absence of SDS. Two-way RM ANOVA. Factor: Timepoint, $F(1,8) = 1.702$, $p = 0.2283$. Factor: Context, $F(1,8) = 88.30$, $p < 0.0001$. Factor: Timepoint x Context, $F(1,8) = 0.3995$, $p = 0.5450$. Post-hoc Tukey ***$p < 0.001$. $n = 9$ mice per group.

0.05. All data are presented as mean ± SEM with overlaid values for individual mice. Details of statistical analyses are found in figure legends.

RESULTS

Stress induces generalization in male and female mice

In the absence of stress (pre-stress retrieval, Fig. 1A), male mice exhibited significantly more freezing in Cxt A compared to in Cxt B (Fig. 1B), indicating weak generalization. After SDS, however, mice no longer showed significant differences in freezing between Cxt A and Cxt B due to significantly higher freezing in Cxt B (Fig. 1B). Accordingly, the generalization index was significantly higher post-stress compared to pre-stress (Fig. 1C). To control for the possibility that this effect was due to an overall increase in fear due to stress, rather than interaction between stress and memory systems, we repeated the same experiment without delivery of footshocks in the conditioning context (Fig. 1D). Under this condition, we found no increase in freezing in Cxt A or Cxt B after SDS, suggesting that SDS alone does not increase freezing (Fig. 1E). Additionally, to control for potential generalization of memories over time, we repeated the same experiment with a 4-day waiting period instead of SDS (Fig. 1F) and also found no increase in freezing in Cxt B (Fig. 1G). Finally, we found that pre-stress retrieval tests were not necessary for the SIG phenotype to develop (Fig. S1), and the order of Cxt A or Cxt B testing during memory retrieval did not affect SIG (Fig. S2). Together, these results suggest that SIG arises from the effects of stress on memory systems.

Next, we sought to study SIG in female mice. SIS was performed on female mice in lieu of SDS, as male CD-1 mice used as aggressors in SDS do not reliably show aggression towards female mice. Similar to male mice, we found that after SIS (Fig. 2A), female mice exhibited significantly higher freezing in Cxt B and a significantly higher generalization index, as compared to before SIS (Fig. 2B, C). Accordingly, mice no longer had significant differences in freezing between Cxt A and Cxt B after SIS (Fig. 2B). To control for the possibilities that the effect was due to an overall increase in fear or due to generalization of negative memories over time, we repeated the same experiment without delivery of footshocks in Cxt A (Fig. 2D) or with a 11-day waiting period instead of SIS (Fig. 2F). We found no increase in freezing in Cxt A or B after stress under either of these conditions (Fig. 2E, G). As in our previous study, the order of Cxt A or Cxt B testing did not affect SIG (Fig. S2). Altogether, these results show that social stress can be a strong trigger of SIG in both male and female mice.

Stress-induced generalization is not associated with anxiety-like or depressive-like behaviors in male and female mice

SDS and SIS have been shown to induce depression-like and anxiety-like behaviors when applied over 10 days and 28 days, respectively [46–54]. We therefore examined whether these behaviors could be related to SIG, which was induced by subacute versions of these stressors in this study. In both male and female mice, we found no significant differences between stressed and non-stressed mice (Figs. S3 and S4), in their latency to emerge in the LDE paradigm or sucrose preference in the SPT. We also found no significant differences in time spent in the open arms in the EPM or time spent in the center region in the OFT, despite the trends towards significance in these two tests in female mice (Fig. S4). Altogether, these results demonstrate that SIG is not associated with anxiety-like and depression-like behaviors that are induced by chronic SDS and SIS.

vGlut2-expressing VTA projections to DH contribute to stress-induced generalization

We recently showed that vGlut2 + VTA afferents to the DH provide a negative valence signal that is sufficient to trigger fear reinstatement after extinction [33]. We therefore examined

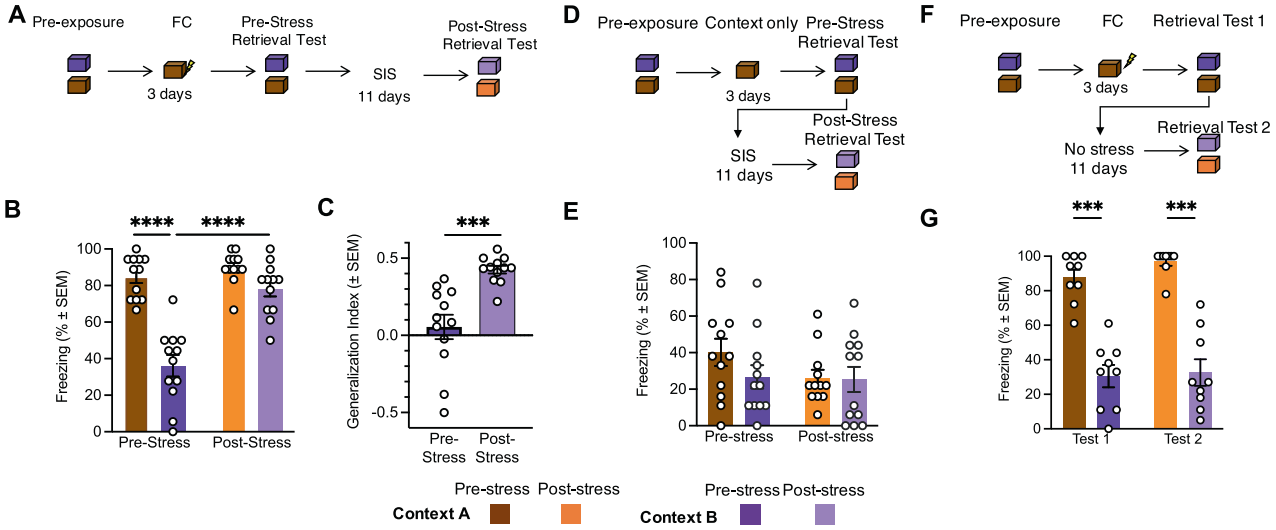


Fig. 2 Stress induces generalization of negative memories in female mice. A Behavioral schematic for SIG performed in female mice and shown in **B**, **C**. **B** Freezing in Context A and Context B in female mice before and after SIS. Two-way RM ANOVA. Factor: Stress, $F(1,11) = 34.81$, $p = 0.0001$. Factor: Context, $F(1,11) = 72.44$, $p < 0.0001$. Factor: Stress x Context, $F(1,11) = 28.66$, $p = 0.0002$. Post-hoc Tukey, $^{****}p < 0.0001$. $n = 12$ mice per group. **C** Generalization index (see methods for formula) in female mice. Wilcoxon matched-pairs signed rank test, $^{***}p = 0.0010$. $n = 12$ mice per group. **D** Behavioral schematic for **E**. Note that mice were exposed to context without footshock. **E** Freezing in Context A and Context B after SIS in the absence of FC. Two-way RM ANOVA. Factor: Stress, $F(1,9) = 0.2024$, $p = 0.6634$. Factor: Context, $F(1,9) = 1.802$, $p = 0.2124$. Factor: Stress x Context, $F(1,9) = 0.006574$, $p = 0.9372$. $n = 10$ mice per group. **F** Behavioral schematic for **G**. Note that mice were exposed to FC but not SIS. **G** Freezing in Context A and Context B after FC in the absence of SIS. Two-way RM ANOVA. Factor: Timepoint, $F(1,8) = 1.776$, $p = 0.2193$. Factor: Context, $F(1,8) = 285.1$, $p < 0.0001$. Factor: Timepoint x Context, $F(1,8) = 2.570$, $p = 0.1476$. Post-hoc Tukey $^{****}p < 0.001$. $n = 9$ mice per group.

whether these projections, by providing a negative valence signal during retrieval of Context B, might also contribute to SIG. In agreement with our previous work [33], GFP-positive terminals in DH were observed after injection of AAV8-hSyn-DIO-GFP into the VTA of vGlut2-Cre mice (Fig. 3A, B). To determine the role of these projections in SIG, we infused AAV8-hSyn-DIO-hM4Di-mCherry into the VTA of vGlut2-Cre male and female mice. Both male (Fig. 3E, F) and female mice (Fig. 3G, H) that received infusion of CNO into DH after stress and before memory retrieval showed significantly reduced generalization indices and Cxt B freezing compared to mice that received infusion of Veh (Fig. 3E–H). Additionally, CNO-infused mice showed significant differences between freezing in Cxt A and Cxt B, whereas Veh-infused mice did not (Fig. 3E, G for males and females, respectively). To rule out the possibility that this effect was due to off-target effects of CNO in the DH, we infused control virus (AAV8-hSyn-DIO-mCherry) into the DH of vGlut1 + mice, as vGlut1 is the predominant glutamatergic neuron subtype in the DH. In these mice, we found no differences in Cxt A or B freezing in mice that received CNO vs. Veh (Fig. S5), suggesting that SIG was not caused by CNO in the absence of hM4Di. To explore the role of these projections in generalization in the absence of stress, male and female mice underwent three days of fear conditioning but were not exposed to social stress afterwards. CNO or Veh was infused before memory retrieval tests. In this case, CNO-infused male and female mice showed no differences in freezing in Cxt A or Cxt B and showed no difference in generalization index compared to Veh-infused mice (Fig. S6). Altogether, these results suggest that VTA vGlut2 + projections to the DH contribute to SIG.

vGlut1-expressing DH projections to RSC contribute to stress-induced generalization

vGlut1 + and vGlut2 + DH projections to RSC are the predominant hippocampal efferents to the neocortex required for the formation (vGlut1 + and vGlut2 +) and retrieval (vGlut1 +) of context memories [39]. We next determined whether these projections also participate in SIG. First, we infused AAV8-hSyn-DIO-mCherry into DH and observed their projections in ventral RSC in vGlut1-Cre (Fig. 4A) and vGlut2-Cre mice (Fig. S7). In agreement with our previous work [39], we found robust mCherry + axon terminals in ventral RSC (Fig. 4B and Fig. S7). Next, we infused AAV8-hSyn-DIO-hM4Di-mCherry into the DH of vGlut1-Cre and vGlut2-Cre mice. After stress, we found that, in both male and female vGlut1-Cre mice (Fig. 4F–H), but not vGlut2-Cre male mice (Fig. S7), infusion of CNO into the RSC before memory retrieval tests led to a significant reduction in the generalization index and Cxt B freezing compared to the Veh-infused group. Additionally, in both male and female vGlut1-Cre mice (Fig. 4E, G), infusion of CNO before memory retrieval tests led to a significant difference in freezing between Cxt A and Cxt B after stress that did not occur in the Veh-infused group. Such effects were not seen in vGlut2-Cre mice (Fig. S7). These results suggest that SIG-relevant information was also conveyed from DH to RSC through vGlut1 + glutamatergic projections.

DISCUSSION

Here, we present a preclinical model for studying SIG by exposing mice to subacute social stress after training them to discriminate aversive from neutral contexts. While we cannot directly establish sex-specific effects due to differences in the social stressors used, it is noteworthy that despite this limitation, different social stressors similarly engaged the extended DH circuit in generating SIG. Furthermore, SIG was not associated with anxiety-like and depressive-like phenotypes that are typically reported after longer exposures to SDS or SIS [46–54]. This suggests that SIG is a separate phenotype and raises an interesting possibility that SIG may be an early cognitive change that ultimately contributes to increased negative affective states.

Memory generalization is often viewed as a problem of memory encoding [23–26], occurring either due to the inability to form sufficiently non-overlapping representations of contexts (e.g., Context A and B) [56], or due to the predominant encoding of general rather than specific features of individual contexts [24].

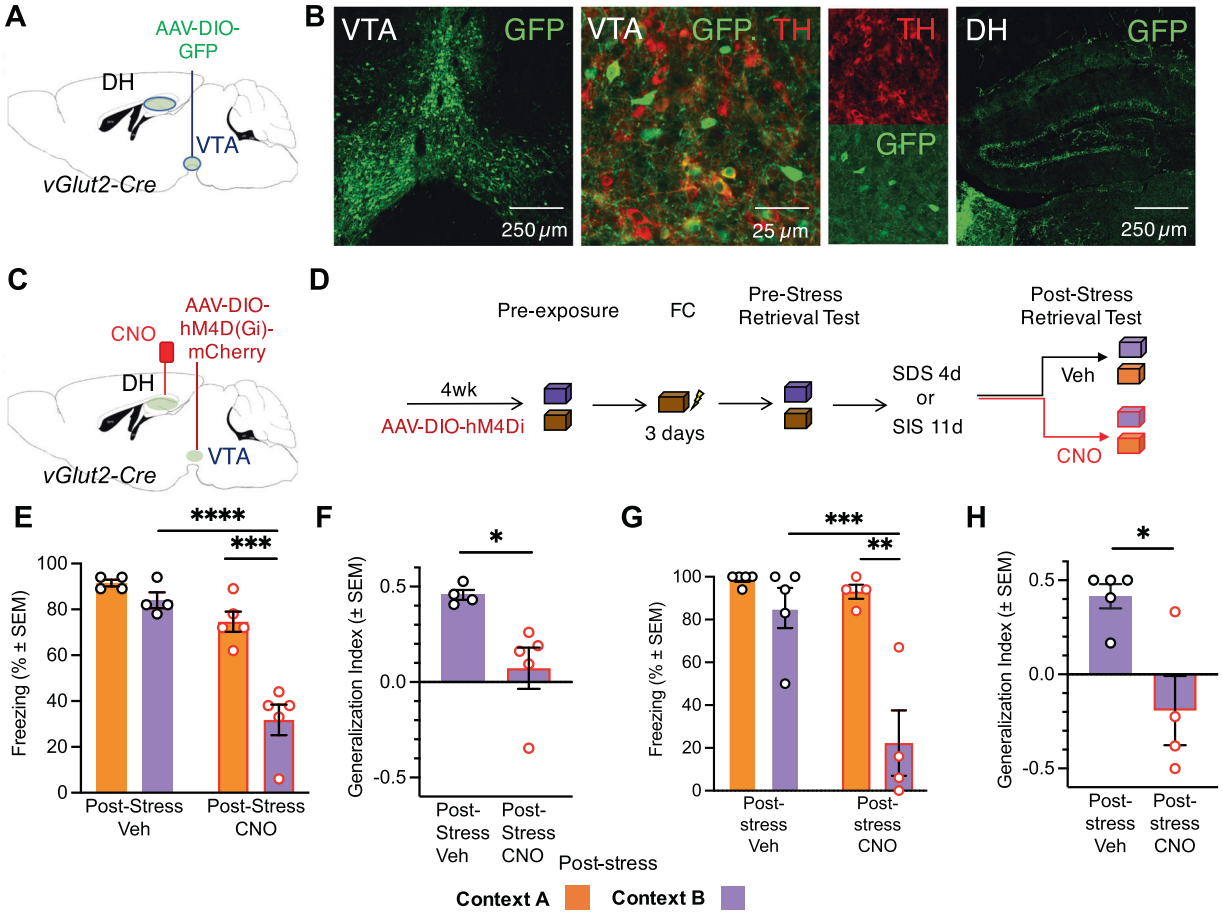


Fig. 3 Inhibition of vGlut2 + VTA projections to DH reduces stress-induced generalization in male and female mice. A Schematic of AAV-DIO-GFP injection in the VTA of vGlut2-Cre mice. **B** Cell bodies of GFP + neurons in VTA and TH immunohistochemistry (left 3 panels) and cell bodies of GFP + neurons in DH (rightmost panel). **C** Schematic of AAV-DIO-hM4D(Gi)-mCherry injection in vGlut2-Cre mice. **D** Behavioral schematic for inhibition of vGlut2 + VTA- >DH projections in **E**, **H**. **E** Freezing in Context A and Context B in SDS-exposed male mice injected with Veh or CNO into the DH before memory retrieval. Two-way RM ANOVA. Factor: Treatment, $F(1,7) = 44.26$, $p = 0.0003$. Factor: Context, $F(1,7) = 32.05$, $p = 0.0008$. Factor: Treatment x Context, $F(1,7) = 15.79$, $p = 0.0054$. Post-hoc Sidak, $^{***}p < 0.001$. $^{****}p < 0.0001$. $n = 4$ mice in Veh group, 5 mice in CNO group. **F** Generalization index in SDS-exposed male mice injected with Veh or CNO into the DH before memory retrieval. Mann–Whitney Test, Sum of Ranks: 30, 15 (Veh, CNO). $^{*}p = 0.0159$. $n = 4$ mice in Veh group, 5 mice in CNO group. **G** Freezing in Context A and Context B in SIS-exposed female mice injected with Veh or CNO into the DH before memory retrieval. Two-way RM ANOVA. Factor: Treatment, $F(1,7) = 14.83$, $p = 0.0063$. Factor: Context, $F(1,7) = 24.73$, $p = 0.0016$. Factor: Treatment x Context, $F(1,7) = 11.49$, $p = 0.0116$. Post-hoc Sidak, $^{**}p < 0.01$, $^{***}p < 0.001$. $n = 5$ mice in Veh group, 4 mice in CNO group. **H** Generalization index in SIS-exposed female mice injected with Veh or CNO into the DH before memory retrieval. Mann–Whitney Test, Sum of Ranks: 34, 11 (Veh, CNO). $^{*}p = 0.0317$. $n = 5$ mice in Veh group, 4 mice in CNO group.

However, such encoding mechanisms are not likely to contribute to SIG, given that several circuit manipulations that we applied at retrieval reversed generalization, thereby revealing that mice were able to encode the contexts in a manner that allowed them to be distinguished from each other. Recent work however, including empirical research and computational models, has emphasized that generalization is also controlled by memory retrieval mechanisms [28, 29]. Our findings show that these mechanisms are highly susceptible to social stress, which may act on the components of the extended hippocampal circuit related to both negative valence [33] and memory [39]. These findings are based on usage of chemogenetic circuit manipulations that we previously extensively validated for their ability to silence both neurons [57] and terminals [33, 39] using slice electrophysiology.

The densest DH glutamatergic projections from VTA terminate in the dentate gyrus and CA2/CA3. There are several mechanisms by which these afferents could contribute to generalization. Based on our recent work [33], we posit that VTA projections to DH provide a negative valence signal that results in reinterpretation of aspects of Context B at retrieval. At the cellular level, this can occur by reactivation of cell ensembles that encode general memory features [24] or by activation of CA3 recurrent loops that support memory generalization [58–60] by VTA afferents, or both.

DH glutamatergic projections to RSC primarily stem from the subiculum. Thus, it is likely that there are one or more intermediary intrahippocampal synapses between VTA glutamatergic projections to DH and DH glutamatergic projections to RSC. Indeed, integration and transmission of information through these microcircuits may then ultimately activate cortical memory representations, which tend to be less specific and more general than hippocampal representations [61, 62]. Therefore, SIG might also target systems consolidation mechanisms mediated by vGlut1-positive DH projections to RSC to bias retrieval toward general rather than specific context features. It should be noted, however, that the effects of our manipulations of vGlut1-positive DH projections to RSC were modest and did not completely diminish SIG, suggesting that targeting vGlut1-positive DH projections to RSC may not be necessary for the full effect of SIG. Consistent with our previous findings [39] in non-stressed mice, vGlut2 + DH- >RSC projections did not significantly

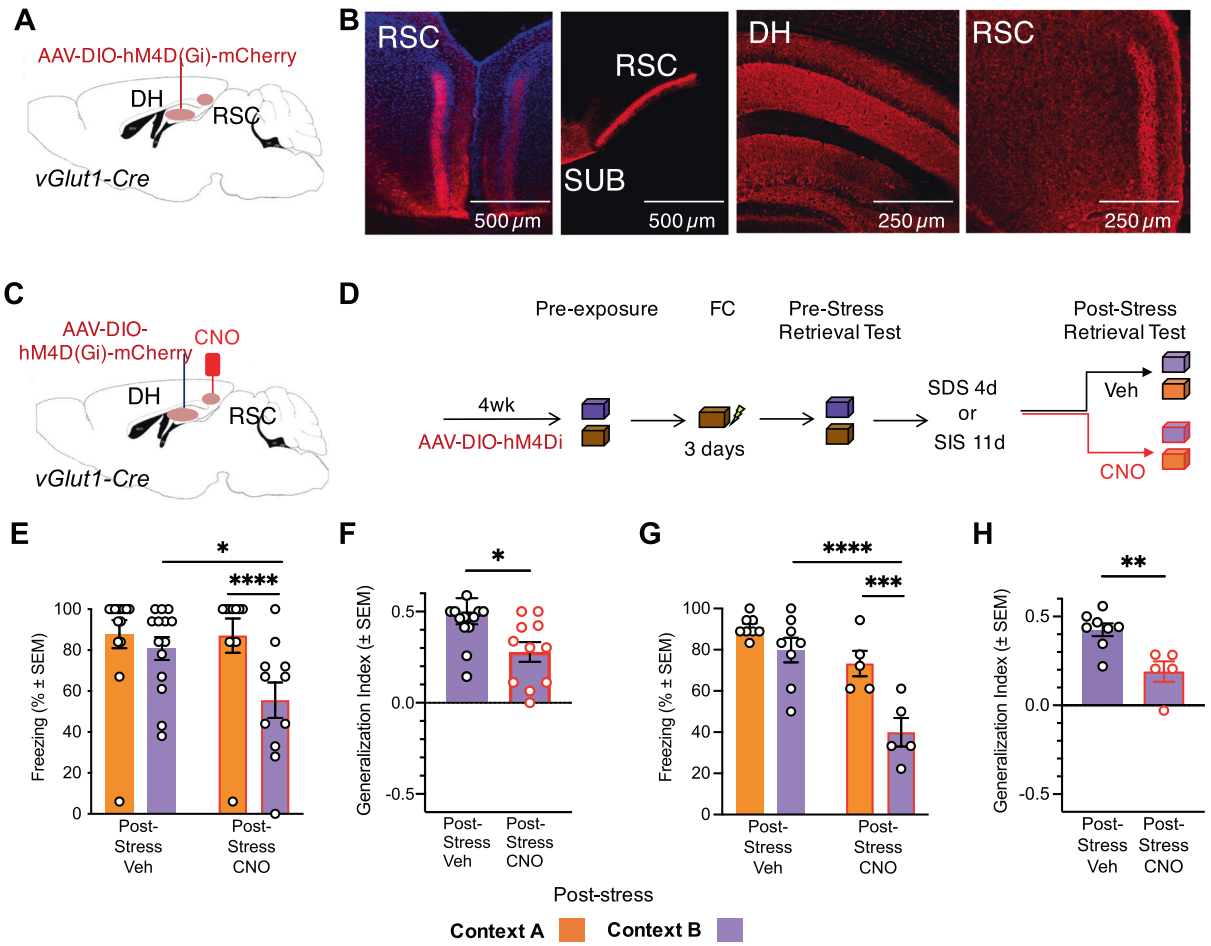


Fig. 4 Inhibition of vGlut1 + DH projections to RSC reduces stress-induced generalization in male and female mice. A Schematic of AAV-hSyn-DIO-mCherry injection in vGlut1-Cre mice. **B** mCherry + neurons in DH and their projections in RSC. **C** Schematic of AAV-DIO-hM4D(Gi)-mCherry injection in vGlut1-Cre mice. **D** Behavioral schematic for inhibition of vGlut1 + DH- >RSC projections during SIG paradigm in **E**–**H**. **E** Freezing in Context A and Context B in SDS-exposed male mice injected with Veh or CNO into the RSC before memory retrieval. Two-way RM ANOVA. Factor: Treatment, $F(1,23) = 1.894$, $p = 0.1820$. Factor: Context, $F(1,23) = 21.41$, $p = 0.0001$. Factor: Treatment x Context, $F(1,23) = 8.681$, $p = 0.0072$. Post-hoc Sidak, $^{*}p < 0.05$, $^{****}p < 0.001$. $n = 14$ mice in Veh group, 11 mice in CNO group. **F** Generalization index in SDS-exposed male mice injected with Veh or CNO into the DH before memory retrieval. Mann–Whitney Test, Sum of Ranks: 226, 99 (Veh, CNO). $^{*}p = 0.0137$. $n = 14$ mice in Veh group, 11 mice in CNO group. **G** Freezing in Context A and Context B in SIS-exposed female mice injected with Veh or CNO into the RSC before memory retrieval. Two-way RM ANOVA. Factor: Treatment, $F(1,11) = 19.24$, $p = 0.0011$. Factor: Context, $F(1,11) = 34.95$, $p = 0.0001$. Factor: Treatment x Context, $F(1,11) = 8.715$, $p = 0.0132$. Post-hoc Sidak, $^{***}p < 0.001$, $^{****}p < 0.001$. $n = 8$ mice in Veh group, 5 mice in CNO group. **H** Generalization index in SIS-exposed female mice injected with Veh or CNO into the RSC before memory retrieval. Mann–Whitney Test, Sum of Ranks: 74, 17 (Veh, CNO). $^{**}p = 0.0062$. $n = 8$ mice in Veh group, 5 mice in CNO group.

contribute to the retrieval of recent context memories either in stressed mice, supporting the specificity of our manipulations and the specificity of the effects of stress on individual projections within the extended hippocampal circuit.

The finding that glutamatergic VTA projections to DH and DH projections to RSC play a role in SIG but not negative memory retrieval is very interesting, especially given that our previous work showed that inhibition of vGlut1 + DH to RSC projections before memory retrieval in the conditioning context reduced conditioning context freezing in the absence of stress [39]. This suggests that either stress alters the neurobiology of this circuit so that it is no longer necessary for retrieval of the original memory, or that the stronger training protocol (3 context-shock associations over 3 days) used in this study might lead to a more robust contextual memory that is represented more redundantly in neuronal networks and may thus be less dependent on this circuit than a single context-shock association [63]. Additionally, this more robust fear memory may contribute to a ceiling effect in freezing levels, which may require more widespread manipulations (e.g.,: co-inhibition of several DH afferents or efferents) to observe a behavioral effect. Finally, it cannot be ruled out that mechanisms involved in retrieval of weaker aversive memories are repurposed towards sub-serving generalization during retrieval of strong memories.

Our results are translationally relevant for several reasons. First, we found that SIG is not associated with other anxiety-like and depressive-like behaviors, suggesting that SIG may be a separate cognitive change independent of affective symptoms, much like overgeneralization is a cognitive phenotype in MDD, GAD, and PTSD that can persist after treatment of affective symptoms. This also leads to the interesting possibility that SIG precedes other anxiety-like and depressive-like behaviors in a similar way as overgeneralization precedes and predicts affective symptoms of MDD, GAD, and PTSD [1, 8, 9]. Second, our finding that SIG occurs during memory retrieval may inform retrieval-based therapeutics for overgeneralization, which may be more clinically effective than

encoding-based therapeutics for overgeneralization. This is primarily due to the fact that patients have already encoded many negative memories by the time they present for treatment. Thus, SIG may be a promising model for researching therapeutic strategies for the prevention and the development of MDD, GAD, and PTSD. Finally, the inclusion of female mice in this study is crucial as MDD, GAD, and PTSD affect significantly more women than men in the United States as well as globally [64–67]. Altogether, these results provide a rodent model to study SIG, elucidate a neural circuit mediating SIG, and provide a framework for future therapeutics targeting overgeneralization.

FUNDING AND DISCLOSURES

This work was supported by National Institute of Mental Health (NIMH) grants R01MH108837 and R01MH078064 awarded to JR and NIMH fellowship F30MH122130 and NIMH Neurobiology of Information Storage training grant T32MH067564 awarded to LYR. The authors have nothing to disclose.

REFERENCES

1. McIntyre RS, Xiao HX, Syeda K, Vinberg M, Carvalho AF, Mansur RB, et al. The prevalence, measurement, and treatment of the cognitive dimension/domain in major depressive disorder. CNS Drugs. 2015;29:577–89.
2. Hitchcock C, Werner-Seidler A, Blackwell SE, Dalgleish T. Autobiographical episodic memory-based training for the treatment of mood, anxiety and stress-related disorders: A systematic review and meta-analysis. Clin Psychol Rev. 2017;52:92–107.
3. Sumner JA, Griffith JW, Mineka S. Overgeneral autobiographical memory as a predictor of the course of depression: a meta-analysis. Behav Res Ther. 2010;48:614–25.
4. Brandes D, Ben-Schachar G, Gilboa A, Bonne O, Freedman S, Shalev AY. PTSD symptoms and cognitive performance in recent trauma survivors. Psychiatry Res. 2002;110:231–8.
5. Reinecke A, Rinck M, Becker ES, Hoyer J. Cognitive-behavior therapy resolves implicit fear associations in generalized anxiety disorder. Behav Res Ther. 2013;51:15–23.
6. Reinecke A, Hoyer J, Rinck M, Becker ES. Cognitive-behavioural therapy reduces unwanted thought intrusions in generalized anxiety disorder. J Behav Ther Exp Psychiatry. 2013;44:1–6.
7. McWhirter L, Ritchie C, Stone J, Carson A. Functional cognitive disorders: a systematic review. Lancet Psychiatry. 2020;7:191–207.
8. Reinecke A, Waldenmaier L, Cooper MJ, Harmer CJ. Changes in automatic threat processing precede and predict clinical changes with exposure-based cognitive-behavior therapy for panic disorder. Biol Psychiatry. 2013;73:1064–70.
9. Jak AJ, Crocker LD, Aupperle RL, Clausen A, Bomyea J. Neurocognition in PTSD: Treatment Insights and Implications. In: Vermetten E, Baker DG, Risbrough VB, editors. Behavioral Neurobiology of PTSD. Cham: Springer International Publishing; 2018. p. 93–116.
10. Reinecke A, Becker ES, Hoyer J, Rinck M. Generalized implicit fear associations in generalized anxiety disorder. Depress Anxiety. 2010;27:252–9.
11. Levy-Gigi E, Szabo C, Richter-Levin G, Keri S. Reduced hippocampal volume is associated with overgeneralization of negative context in individuals with PTSD. Neuropsychology. 2015;29:151–61.
12. Carver CS. Generalization, adverse events, and development of depressive symptoms. J Pers. 1998;66:607–19.
13. Ganellen RJ. Specificity of attributions and overgeneralization in depression and anxiety. J Abnorm Psychol. 1988;97:83–6.
14. Thome J, Hauschild S, Koppe G, Liebke L, Rausch S, Herzog JI, et al. Generalisation of fear in PTSD related to prolonged childhood maltreatment: an experimental study. Psychol Med. 2018;48:2223–34.
15. Lenaert B, Boddez Y, Vervliet B, Schruers K, Hermans D. Reduced autobiographical memory specificity is associated with impaired discrimination learning in anxiety disorder patients. Front Psychol. 2015;6:889.
16. Lissek S, Kaczkurkin AN, Rabin S, Geraci M, Pine DS, Grillon C. Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. Biol Psychiatry. 2014;75:909–15.
17. Kaczkurkin AN, Burton PC, Chazin SM, Manbeck AB, Espensen-Sturges T, Cooper SE, et al. Neural substrates of overgeneralized conditioned fear in PTSD. Am J Psychiatry. 2017;174:125–34.
18. Morey RA, Dunsmoor JE, Haswell CC, Brown VM, Vora A, Weiner J, et al. Fear learning circuitry is biased toward generalization of fear associations in posttraumatic stress disorder. Transl Psychiatry. 2015;5:e700.
19. Beck ATRA, Shaw BF, Emery G. Cognitive therapy of depression. Guilford: New York; 1979.
20. Fulford D, Rosen RK, Johnson SL, Carver CS. Negative generalization and symptoms of anxiety disorders. J Exp Psychopathol. 2012;3:62–8.
21. Besnard A, Sahay A. Adult hippocampal neurogenesis, fear generalization, and stress. Neuropsychopharmacology. 2016;41:24–44.
22. Dunsmoor JE, Otto AR, Phelps EA. Stress promotes generalization of older but not recent threat memories. Proc Natl Acad Sci USA. 2017;114:9218–23.
23. Shohamy D, Turk-Browne NB. Mechanisms for widespread hippocampal involvement in cognition. J Exp Psychol Gen. 2013;142:1159–70.
24. Sun X, Bernstein MJ, Meng M, Rao S, Sorensen AT, Yao L, et al. Functionally distinct neuronal ensembles within the memory engram. Cell. 2020;181:410–23 e17.
25. Frankland PW, Cestari V, Filipkowski RK, McDonald RJ, Silva AJ. The dorsal hippocampus is essential for context discrimination but not for contextual conditioning. Behav Neurosci. 1998;112:863–74.
26. Wiltgen BJ, Silva AJ. Memory for context becomes less specific with time. Learn Mem. 2007;14:313–7.
27. Abdou K, Shehata M, Choko K, Nishizono H, Matsuo M, Muramatsu SI, et al. Synapse-specific representation of the identity of overlapping memory engrams. Science. 2018;360:1227–31.
28. Kumaran D, McClelland JL. Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. Psychol Rev. 2012;119:573–616.
29. Banino A, Koster R, Hassabis D, Kumaran D. Retrieval-based model accounts for striking profile of episodic memory and generalization. Sci Rep. 2016;6:31330.
30. Leutgeb JK, Leutgeb S, Moser MB, Moser EI. Pattern separation in the dentate gyrus and CA3 of the hippocampus. Science. 2007;315:961–6.
31. Treves A, Tashiro A, Witter MP, Moser EI. What is the mammalian dentate gyrus good for? Neuroscience. 2008;154:1155–72.
32. Rolls ET, Kesner RP. Pattern separation and pattern completion in the hippocampal system. Introduction to the Special Issue. Neurobiol Learn Mem. 2016;129:1–3.
33. Han Y, Zhang Y, Kim H, Grayson VS, Jovasevic V, Ren W, et al. Excitatory VTA to DH projections provide a valence signal to memory circuits. Nat Commun. 2020;11:1466.
34. Roy DS, Kitamura T, Okuyama T, Ogawa SK, Sun C, Obata Y, et al. Distinct Neural circuits for the formation and retrieval of episodic memories. Cell. 2017;170:1000–12 e19.
35. Stern CE, Corkin S, Gonzalez RG, Guimaraes AR, Baker JR, Jennings PJ, et al. The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. Proc Natl Acad Sci USA. 1996;93:8660–5.
36. Wild JM, Blampied NM. Hippocampal lesions and stimulus generalization in rats. Physiol Behav. 1972;9:505–11.
37. Bucci DJ, Saddoris MP, Burwell RD. Contextual fear discrimination is impaired by damage to the postrhinal or perirhinal cortex. Behav Neurosci. 2002;116:479–88.
38. Rudy JW, O'Reilly RC. Contextual fear conditioning, conjunctive representations, pattern completion, and the hippocampus. Behav Neurosci. 1999;113:867–80.
39. Yamawaki N, Corcoran KA, Guedea AL, Shepherd GMG, Radulovic J. Differential contributions of glutamatergic hippocampal→retrosplenial cortical projections to the formation and persistence of context memories. Cereb Cortex. 2018; 29:2728–36.
40. Jiang MY, DeAngeli NE, Bucci DJ, Todd TP. Retrosplenial cortex has a time-dependent role in memory for visual stimuli. Behav Neurosci. 2018;132:396–402.
41. Todd TP, Mehlman ML, Keene CS, DeAngeli NE, Bucci DJ. Retrosplenial cortex is required for the retrieval of remote memory for auditory cues. Learn Mem. 2016;23:278–88.
42. Corcoran KA, Frick BJ, Radulovic J, Kay LM. Analysis of coherent activity between retrosplenial cortex, hippocampus, thalamus, and anterior cingulate cortex during retrieval of recent and remote context fear memory. Neurobiol Learn Mem. 2016;127:93–101.
43. Corcoran KA, Leaderbrand K, Radulovic J. Extinction of remotely acquired fear depends on an inhibitory NR2B/PKA pathway in the retrosplenial cortex. J Neurosci. 2013;33:19492–8.
44. Leaderbrand K, Chen HJ, Corcoran KA, Guedea AL, Jovasevic V, Wess J, et al. Muscarinic acetylcholine receptors act in synergy to facilitate learning and memory. Learn Mem. 2016;23:631–38.
45. Corcoran KA, Donnan MD, Tronson NC, Guzman YF, Gao C, Jovasevic V, et al. NMDA receptors in retrosplenial cortex are necessary for retrieval of recent and remote context fear memory. J Neurosci. 2011;31:11655–9.
46. Kudryavtseva NN, Bakshtanovskaya IV, Koryakina LA. Social model of depression in mice of C57BL/6J strain. Pharm Biochem Behav. 1991;38:315–20.
47. Rygula R, Abumaria N, Flugge G, Fuchs E, Ruther E, Havemann-Reinecke U. Anhedonia and motivational deficits in rats: impact of chronic social stress. Behav Brain Res. 2005;162:127–34.
48. Rygula R, Abumaria N, Flugge G, Hiemke C, Fuchs E, Ruther E, et al. Citalopram counteracts depressive-like symptoms evoked by chronic social stress in rats. Behav Pharmacol. 2006;17:19–29.
49. Rygula R, Abumaria N, Domenici E, Hiemke C, Fuchs E. Effects of fluoxetine on behavioral deficits evoked by chronic social stress in rats. Behav Brain Res. 2006;174:188–92.

50. Berton O, McClung CA, Dileone RJ, Krishnan V, Renthal W, Russo SJ, et al. Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress. Science. 2006;311:864–8.
51. Golden SA, Covington HE 3rd, Berton O, Russo SJ. A standardized protocol for repeated social defeat stress in mice. Nat Protoc. 2011;6:1183–91.
52. Labaka A, Gomez-Lazaro E, Vegas O, Perez-Tejada J, Arregi A, Garmendia L. Reduced hippocampal IL-10 expression, altered monoaminergic activity and anxiety and depressive-like behavior in female mice subjected to chronic social instability stress. Behav Brain Res. 2017;335:8–18.
53. Goni-Balentziaga O, Perez-Tejada J, Renteria-Dominguez A, Lebena A, Labaka A. Social instability in female rodents as a model of stress related disorders: A systematic review. Physiol Behav. 2018;196:190–99.
54. Haller J, Fuchs E, Halasz J, Makara GB. Defeat is a major stressor in males while social instability is stressful mainly in females: towards the development of a social stress model in female rats. Brain Res Bull. 1999;50:33–9.
55. Xu W, Sudhof TC. A neural circuit for memory specificity and generalization. Science. 2013;339:1290–5.
56. Berens SC, Bird CM. The role of the hippocampus in generalizing configural relationships. Hippocampus. 2017;27:223–28.
57. Jovasevic V, Corcoran KA, Leaderbrand K, Yamawaki N, Guedea AL, Chen HJ, et al. GABAergic mechanisms regulated by miR-33 encode state-dependent fear. Nat Neurosci. 2015;18:1265–71.
58. Rolls ET. A theory of hippocampal function in memory. Hippocampus. 1996;6:601–20.
59. Rolls ET. Pattern separation, completion, and categorisation in the hippocampus and neocortex. Neurobiol Learn Mem. 2016;129:4–28.
60. Rolls ET, Kesner RP. A computational theory of hippocampal function, and empirical tests of the theory. Prog Neurobiol. 2006;79:1–48.
61. Wiltgen BJ, Zhou M, Cai Y, Balaji J, Karlsson MG, Parivash SN, et al. The hippocampus plays a selective role in the retrieval of detailed contextual memories. Curr Biol. 2010;20:1336–44.
62. de Sousa AF, Cowansage KK, Zutshi I, Cardozo LM, Yoo EJ, Leutgeb S, et al. Optogenetic reactivation of memory ensembles in the retrosplenial cortex induces systems consolidation. Proc Natl Acad Sci USA. 2019;116:8576–81.
63. Wiltgen BJ, Sanders MJ, Anagnostaras SG, Sage JR, Fanselow MS. Context fear learning in the absence of the hippocampus. J Neurosci. 2006;26:5484–91.
64. Villarroel MA, Terlizzi EP. Symptoms of depression among adults: United States, 2019. NCHS Data Brief. 2020;379:1–8.
65. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 2005;62:593–602.
66. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 2005;62:617–27.
67. Whiteford HA, Ferrari AJ, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. PLoS ONE. 2015;10:e0116820.

ACKNOWLEDGEMENTS

We would like to thank Dr. John Kessler, Dr. Ana Cicvaric, Dr. Hui Zhang, and Dr. Vladimir Jovasevic for helpful discussions regarding the work described here.

AUTHOR CONTRIBUTIONS

LYR and JR designed and LYR, MAAM, VSG and PG conducted the experiments. LYR and ALG bred and genotyped the transgenic mice. LYR analyzed the data. LYR and JR wrote the paper, and MAAM provided significant edits. All authors discussed and commented on the paper.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41386-021-01174-4.

Correspondence and requests for materials should be addressed to Lynn Y. Ren or Jelena Radulovic.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

frontiers in
HUMAN NEUROSCIENCE

REVIEW ARTICLE
published: 09 November 2009
doi: 10.3389/neuro.09.031.2009

The human brain in numbers: a linearly scaled-up primate brain

Suzana Herculano-Houzel*

Instituto de Ciências Biomédicas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil

Edited by:
Andreas Jeromin,
Allen Institute for Brain Science, USA

Reviewed by:
Karl Herrup,
Case Western University, USA
Robert Barton,
University of Durham, UK

****Correspondence:***
Suzana Herculano-Houzel, Laboratório de Neuroanatomia Comparada, Instituto de Ciências Biomédicas, Universidade Federal do Rio de Janeiro, Rua Carlos Chagas Filho 373, 21950-902 Rio de Janeiro, Rio de Janeiro, Brazil.
e-mail: suzanahh@gmail.com

The human brain has often been viewed as outstanding among mammalian brains: the most cognitively able, the largest-than-expected from body size, endowed with an overdeveloped cerebral cortex that represents over 80% of brain mass, and purportedly containing 100 billion neurons and 10× more glial cells. Such uniqueness was seemingly necessary to justify the superior cognitive abilities of humans over larger-brained mammals such as elephants and whales. However, our recent studies using a novel method to determine the cellular composition of the brain of humans and other primates as well as of rodents and insectivores show that, since different cellular scaling rules apply to the brains within these orders, brain size can no longer be considered a proxy for the number of neurons in the brain. These studies also showed that the human brain is not exceptional in its cellular composition, as it was found to contain as many neuronal and non-neuronal cells as would be expected of a primate brain of its size. Additionally, the so-called overdeveloped human cerebral cortex holds only 19% of all brain neurons, a fraction that is similar to that found in other mammals. In what regards absolute numbers of neurons, however, the human brain does have two advantages compared to other mammalian brains: compared to rodents, and probably to whales and elephants as well, it is built according to the very economical, space-saving scaling rules that apply to other primates; and, among economically built primate brains, it is the largest, hence containing the most neurons. These findings argue in favor of a view of cognitive abilities that is centered on absolute numbers of neurons, rather than on body size or encephalization, and call for a re-examination of several concepts related to the exceptionality of the human brain.

Keywords: brain scaling, number of neurons, human, encephalization

INTRODUCTION

THE HUMAN BRAIN AS A SPECIAL BRAIN

What makes us human? Is our brain, the only one known to study other brains, special in any way? According to a recent popular account of what makes us unique, "we have brains that are bigger than expected for an ape, we have a neocortex that is three times bigger than predicted for our body size, we have some areas of the neocortex and the cerebellum that are larger than expected, we have more white matter" – and the list goes on (Gazzaniga, 2008). Most specialists seem to agree (for example, Marino, 1998; Rilling, 2006; Sherwood et al., 2006). Since ours is obviously not the largest brain on Earth, our superior cognitive abilities cannot be accounted for by something as simple as brain size, the most readily measurable parameter regarding the brain. Emphasis is thus placed on an exceptionality that is, curiously, not brain-centered, but rather body-centered: With a smaller body but a larger brain than great apes, the human species deviates from the relationship between body and brain size that applies to other primates, great apes included, boasting a brain that is 5–7× too large for its body size (Jerison, 1973; Marino, 1998). Recent efforts to support this uniqueness have focused on finding genetic differences between humans and other primates (reviewed in Vallender, 2008), as well as cellular particularities such as the presence and distribution of Von Economo neurons (Nimchinsky et al., 1999; but see Butti et al., 2009; Hakeem et al., 2009).

To regard the human brain as unique requires considering it to be an outlier: an exception to the rule, whatever that rule is. This makes little sense, however, in light of evolution. If we go to such great lengths to affirm, and teach, that evolution is the origin of diversity in life, and to find trends and laws that apply to kingdoms, phyla and orders as a whole, why then insist that whatever scaling rules apply to other primates must not apply to us? In view of the vexing size inferiority in brain size and of the lack of information about what our brains are actually made of – and how that compares to other brains, particularly those of whales and elephants – resorting to a quest for uniqueness may have seemed as a necessary, natural step to justify the cognitive superiority of the human brain.

Recently, a novel quantitative tool developed in our lab (Herculano-Houzel and Lent, 2005) has finally made the numbers of neurons and non-neuronal cells that compose the brains of various mammals, humans included, available for comparative analysis. This review will focus on such a quantitative, comparative analysis, with emphasis on the numbers that characterize the human brain: what they are, how they have been viewed in the past, and how they change our view of where the human brain fits into the diversity of the mammalian nervous system.

THE HUMAN BRAIN IN NUMBERS

How many neurons does the human brain have, and how does that compare to other species? Many original articles, reviews and textbooks affirm that we have 100 billion neurons and 10 times more glial cells (Kandel et al., 2000; Ullian et al., 2001; Doetsch, 2003; Nishiyama et al., 2005; Noctor et al., 2007; Allen and Barres, 2009), usually with no references cited. This leaves the reader with

the impression that the cellular composition of the human brain has long been determined. Indeed, an informal survey with senior neuroscientists that we ran in 2007 showed that most believed that the number of cells in the human brain was indeed already known: that we have about 100 billion neurons, outnumbered by about 10 times more glial cells – but none of the consulted scientists could cite an original reference for these numbers (Herculano-Houzel and Lent, unpublished observations). Curiously, the widespread concept that neurons represent about 10% of all cells in the human brain might be one of the arguments behind the popular, but mistaken, notion that we only use 10% of our brain (Herculano-Houzel, 2002).

The reason for such lack of references is that indeed there was, to our knowledge, no actual, direct estimate of numbers of cells or of neurons in the entire human brain to be cited until 2009. A reasonable approximation was provided by Williams and Herrup (1988), from the compilation of partial numbers in the literature. These authors estimated the number of neurons in the human brain at about 85 billion: 12–15 billion in the telencephalon (Shariff, 1953), 70 billion in the cerebellum, as granule cells (based on Lange, 1975), and fewer than 1 billion in the brainstem. With more recent estimates of 21–26 billion neurons in the cerebral cortex (Pelvig et al., 2008) and 101 billion neurons in the cerebellum (Andersen et al., 1992), however, the total number of neurons in the human brain would increase to over 120 billion neurons.

As to the 10 times more numerous glial cells in the human brain, that seems to be the case only in subcortical nuclei such as the thalamus (17 glial cells per neuron) and the ventral pallidum (12 glial cells per neuron; Pakkenberg and Gundersen, 1988). In the gray matter of the cerebral cortex, glial cells outnumber neurons by a factor of <2 (Sherwood et al., 2006; Pelvig et al., 2008). Given the relatively small number of glial cells reported for the human cerebellum, where they are outnumbered by neurons by at least 25:1 (Andersen et al., 1992), the only possible explanation for the ubiquitous quote of 10 times more glial than neuronal cells in the entire human brain would be the presence of nearly one trillion glial cells in the remaining structures alone – an unlikely scenario, since these structures represent <10% of total brain mass.

WHY BOTHER WITH CELL NUMBERS?

Across species, the number of neurons and their relative abundance in different parts of the brain is widely considered to be a determinant of neural function and, consequently, of behavior (Williams and Herrup, 1988). Among mammals, those species with the largest brains, such as cetaceans and primates, have a greater range and versatility of behavior than those with the smallest brains, such as insectivores (Jerison, 1985; Marino, 2002). Among birds, those that are larger-brained (corvids, parrots and owls) are also considered the most intelligent (Lefebvre et al., 2004). A recent comparison of several parameters, including brain size, relative brain size, encephalization, conduction velocity and estimated numbers of neurons led two authors to conclude that the "factors that correlate better with intelligence (across species) are the number of cortical neurons and conduction velocity, as the basis for information processing" (Roth and Dicke, 2005). Indeed, within non-human primates, a recent meta-analysis concluded that the best predictor of the cognitive abilities of a species is absolute brain size, not relative size nor encephalization quotient (EQ; Deaner et al., 2007).

However, the correlation between absolute brain size and cognitive abilities breaks down when species of similar brain size are compared across orders. Monkeys, for instance, possess brains that are much smaller than those of ungulates, but the higher cognitive and behavioral flexibility of monkeys over ungulates is anecdotally evident to any observer who compares the ingenious and complex abilities of macaques to those of cows or horses, even though the latter have 4–5× larger brains than macaques. For similar-sized brains, rodents also perform more poorly than primates: With a brain of only 52 g, the behavioral, social and cognitive repertoire of the capuchin monkey is outstanding compared to the capybara, a giant Amazonian rodent (MacDonald, 1981), even though the latter has a larger brain of 75 g. This is reminiscent of the most striking and troubling discrepancy regarding brain size and cognitive abilities: that between humans and larger-brained species such as whales and elephants. If the latter have brains that are up to six times larger than a human brain, why should we be more cognitively able? Answering this question requires a direct examination of the numbers of neurons that compose the brains of humans and other species.

BRAIN AND BODY SCALING: THE TRADITIONAL VIEW

ASSUMPTION 1: BODY SIZE MATTERS

If the smaller size of the human brain compared to elephant and whale brains (**Figure 1**) translates into a smaller number of neurons in the human brain than in the latter, then what makes the human brain outstanding in its cognitive abilities? In the absence of direct estimates of numbers of neurons in these and other species, the search for a neural correlate for human capacities has placed emphasis on the characteristic that most undisputedly places humans above other mammals: the EQ (Jerison, 1973). This measure is based on the observation that, across species, brain size correlates with body size in a way that can be described mathematically with a power function, thus allowing the predicted brain mass to be calculated for any species. EQ indicates how much the observed brain mass of a species deviates from the expected for its body mass: an EQ of 1 indicates that the observed brain mass matches the expected value; an EQ >1 means that brain size in that species is larger than expected for its body mass.

Compared to mammals as a whole, humans have the largest EQ found for any mammal, of between 7 and 8 (Jerison, 1973); even if compared to anthropoid primates only, humans still have an EQ of over 3, a value that is larger than that obtained for any other

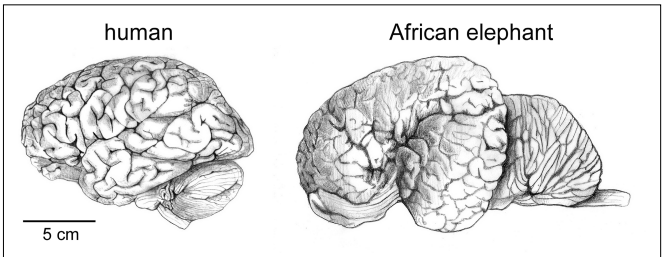


FIGURE 1 |The human brain is not the largest. Brains of a human and of an African elephant are depicted here at the same scale. Drawings by Lorena Kaz based on images freely available from the University of Wisconsin and Michigan State Comparative Mammalian Brain Collections (www.brainmuseum.org).

primate or cetacean (Marino, 1998). The position of the human species as an outlier in the body × brain comparison is made clear if one considers that although gorillas and orangutans overlap or exceed humans in body size, their brains amount to only about one-third of the size of the human brain.

There are, however, several problems with the notion that the explanation for the superior cognitive abilities of the human species lies in its large EQ. For one, it is not obvious how larger-than-expected brain mass would confer a cognitive advantage. In principle, this advantage would rely on the availability for cognitive functions of whatever brain mass exceeds what is necessary to process body-related information. However, according to this notion, small-brained animals with very large EQs should be expected to have more cognitive abilities than large-brained animals with smaller EQs. Capuchin monkeys, for instance, have much larger EQs than gorillas (Marino, 1998), but are outranked by these in cognitive performance (Deaner et al., 2007). Absolute brain mass and number of neurons, left out of the encephalization equation, must clearly be taken into consideration, since the "exceeding number of neurons" in a large brain should necessarily be larger than that in a smaller brain of same EQ (Herculano-Houzel, 2007).

Another problem with the utility of the EQ is that the body–brain mass relationship from which expected brain mass is derived depends on the precise combination of species computed (Barton, 2006; Herculano-Houzel et al., 2007). We have recently found that, compared to the linear brain × body relationship that applies to the primate species in our sample (which consisted of simian and prosimian primates; Herculano-Houzel et al., 2007), the human brain deviates by only 10% from its expected size (Azevedo et al., 2009). This conformity to the body × brain relationship that applies to non-anthropoid primates is consistent with the observation that, like in other non-anthropoid primates, the human brain mass represents about 2% of body mass. Given the sensitivity of EQ to the species included and our finding that the human brain conforms to the scaling rules that apply to other primates (see below), we have suggested that, rather than humans having a larger brain than expected, it is the great apes such as orangutans and, more notably, gorillas that have bodies that are much larger than expected for primates of their brain size (Herculano-Houzel et al., 2007).

This latter possibility of a dissociation between brain and body development in evolution (which might be only circumstantially, and not causally, related) constitutes a final criticism to the usefulness of the EQ as an index of "brain" evolution in comparative studies: indeed, the emphasis on the body-centered EQ overlooks the observation that, compared to other mammalian orders, primate encephalization is the result of a shift in postcranial growth processes, not a modification of brain growth (Deacon, 1997). In the words of Deacon, "if primates have big brains merely because they have small bodies, we cannot presume that this represents an evolutionary trend driven by cognitive demands" (p. 343). In this scenario, however, the human brain exhibits a further modification in that it continues to grow as though in a larger body (Deacon, 1997).

ASSUMPTION 2: BRAIN SIZE MATTERS

Brain size varies across mammals by a factor of approximately 100,000 (Tower, 1954; Stolzenburg et al., 1989). Different mammalian orders have traditionally been pooled together in studies of brain allometry, as if their brains were built according to the same scaling rules (for example, Haug, 1987; Zhang and Sejnowski, 2000).

Comparisons across orders that seem to invalidate the correlation between numbers of neurons and cognitive ability, such as between monkeys and ungulates, or rodents and primates, also bear this hidden caveat: the assumption that brain size relates to number of neurons in the brain in a similar fashion across orders. This assumption, which was justifiable by the lack of direct estimates of the neuronal composition of the brain of different species, is so widespread that it implicitly or explicitly underlies most comparative studies to date (for example, Haug, 1987; Finlay and Darlington, 1995; Barton and Harvey, 2000; Clark et al., 2001). The very concept of encephalization presupposes that not only the brain scales as a function of body size, but that all brains scale the same way, such that the only informative (and sufficient) variable is brain size and its deviation from the expected. However, our quantitative studies on the cellular scaling rules that apply to different mammalian orders have shown that this assumption is invalid and therefore should no longer be applied (see below).

ASSUMPTION 3: PROPORTIONS AND RELATIVE SIZE MATTER

An often cited argument in favor of the uniqueness of the human brain is its relatively large cerebral cortex, which accounts for 82% of brain mass. Within this large cerebral cortex, a relative enlargement of the prefrontal cortex was once considered a hallmark of the human brain, but this view has however been overthrown by modern measurements (Semendeferi et al., 2002). Still, the distribution of cortical mass in humans may differ from that in other primates, endowing particularly relevant regions such as area 10 with relatively more neurons in the human cortex (Semendeferi et al., 2001).

Relative size is supposed to be a meaningful indicator of relative functional importance of a brain structure based on the assumption that it is a proxy for relative number of neurons. For instance, the increase in relative size of the cerebral cortex with increasing brain size simultaneously with no systematic change in the relative size of the cerebellum has been used as evidence that these structures are functionally independent and have been evolving separately (Clark et al., 2001). Such discrepancy would support the popular notion that brain evolution equates with development of the cerebral cortex, which comes to predominate over the other brain structures. However, analysis of absolute, rather than relative, cerebral cortical and cerebellar volumes in the same dataset leads to the opposite conclusion: the coordinated scaling of these volumes, as well as of the surface areas of these structures, would be evidence that the cerebral cortex and cerebellum are functionally related and have been evolving coordinately (Barton, 2002; Sultan, 2002).

As it turns out, however, the underlying assumption that the relative size of a brain structure reflects the relative number of brain neurons that it contains is flawed.

Now that numbers of neurons are available across rodents, primates and insectivores, we find that the cerebral cortex, despite varying in relative size from 42% (in the mouse) to 82% of brain mass (in the human), contains between 13 and 28% of all brain neurons in 15 of 18 species studied, ranging between 13% (in moles) and 41% (in the squirrel monkey; Herculano-Houzel et al.,

2006, 2007; Sarko et al., 2009). Most importantly, this fractional number of neurons in the cerebral cortex relative to the whole brain is not correlated with the relative size of the cerebral cortex (**Figure 2**). Instead, the number of neurons in the cerebral cortex increases coordinately with the number of neurons in the cerebellum (Herculano-Houzel, submitted).

A NEW VIEW: SCALING OF NEURONAL NUMBERS

CELLULAR SCALING RULES FOR RODENT, INSECTIVORE, AND PRIMATE BRAINS

Our group has been investigating the cellular scaling rules that apply to brain allometry in different mammalian orders using the novel method of isotropic fractionation, which produces cell counts derived from tissue homogenates from anatomically defined brain regions (Herculano-Houzel and Lent, 2005). Through the estimation of absolute numbers of neuronal and non-neuronal cells in the brains of different mammalian species and their comparison within individual orders, we have been able to determine the scaling rules that apply to the brains of species spanning a wide range of body and brain masses in rodents (Herculano-Houzel et al., 2006), primates (Herculano-Houzel et al., 2007) and more recently in insectivores (Sarko et al., 2009). A comparative overview of brain mass and total number of neurons for these species can be seen in **Figure 3**.

A recent issue in comparative studies of brain scaling has been the examination of how residual variation in different parameters relate to phylogenetic relationships once shared evolutionary commonalities in body or brain size are accounted for (Harvey and Pagel, 1991; Nunn and Barton, 2000). Although such analyses of independent contrasts are instrumental for identifying evolutionary correlations across taxa while taking into account this phylogenetic nonindependence, they overlook the very issue at hand here: how the size of the brain reflects the number of neurons that it contains, regardless of body size and of any other shared characteristics. For this reason, the analysis reviewed here, referred to as unveiling the "cellular scaling rules" for the brain of different mammalian orders, considers solely how brain size changes as a function of its number of neurons within a given order, irrespective of body size, and without any concerns regarding phylogenenetic effects within that order, or even whether evolution of the extant species has involved an expansion of brain size, a reduction, or both. In the particular case of primates, we have recently extended our analysis to another set of five primate species (Gabi et al., submitted), and found that the same cellular scaling relationships apply to the original dataset (Herculano-Houzel et al., 2007), to the second dataset, and to the combined, extended dataset. This is evidence that the cellular scaling rules considered here from a set of primate species also extend to primates as a whole, and can be used to infer the expected cellular composition of the human brain – even though small variations may occur across species that might, indeed, be due to phylogenetic interdependencies.

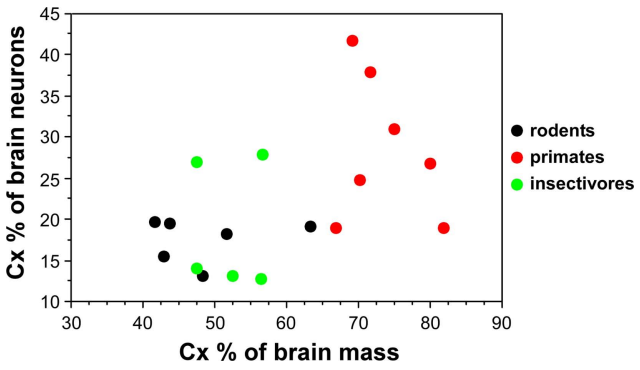


FIGURE 2 | Relative size of the cerebral cortex does not inform about the relative number of neurons in the cortex compared to the whole brain. Each point indicates, for a given species, the average relative cortical mass as a percentage of total brain mass (X-axis) and the average relative number of cortical neurons as a percentage of the total number of neurons in the brain (Y-axis). Data from Herculano-Houzel et al. (2006, 2007); Azevedo et al. (2009); and Sarko et al. (2009).

In the order Rodentia, we find that the brain increases in size faster than it gains neurons, with a decrease in neuronal densities which, in the presence of constant non-neuronal cell densities, implies that average neuronal size increases rapidly as neurons become more numerous (Herculano-Houzel et al., 2006). The increase in numbers of neurons in the cerebral cortex, cerebellum and remaining areas is concurrent with an even greater increase in numbers of non-neurons, yielding a maximal glia/neuron ratio that increases with brain size (Herculano-Houzel et al., 2006). These findings corroborated previous studies describing neuronal density decreasing and the glia-to-neuron ratio increasing with increasing brain size across mammalian taxa (Tower and Elliot, 1952; Shariff, 1953; Friede, 1954; Tower, 1954; Hawking and Olszewski, 1957; Haug, 1987; Reichenbach, 1989; Stolzenburg et al., 1989).

In contrast to rodent brains, which scale hypermetrically in size with their numbers of neurons, primate brain size increases approximately isometrically as a function of neuron number, with no systematic change in neuronal density or in the non-neuronal/neuronal ratio with increasing brain size (Herculano-Houzel et al., 2007). Across insectivore species, on the other hand, the cerebellum increases linearly in size as a function of its number of neurons (as in primates), while the cerebral cortex increases in size hypermetrically as it gains neurons (as in rodents; Sarko et al., 2009). In view of the similar non-neuronal cell densities across species, hypermetric scaling of brain structure mass as a function of its number of neurons implies a concurrent increase in the average neuronal size (which, in the method's definition, includes not only the cell soma but also the entire dendritic and axonal arborizations as well as synapses; Herculano-Houzel et al., 2006). As a consequence of these different cellular scaling rules, shown in **Table 1**, a 10-fold increase in the number of neurons in a rodent brain results in a 35-fold larger brain; in contrast, a similar 10-fold increase in the number of neurons in a primate brain results in an increase in brain size of only 11-fold.

NOT ALL BRAINS ARE CREATED EQUAL: COGNITIVE ABILITIES AND NUMBERS OF NEURONS

The different cellular scaling rules that apply to rodent, primate and insectivore brains show very clearly that brain size cannot be used indiscriminately as a proxy for numbers of neurons in the brain, or even in a brain structure, across orders. By maintaining the average neuronal size (including all arborizations) invariant

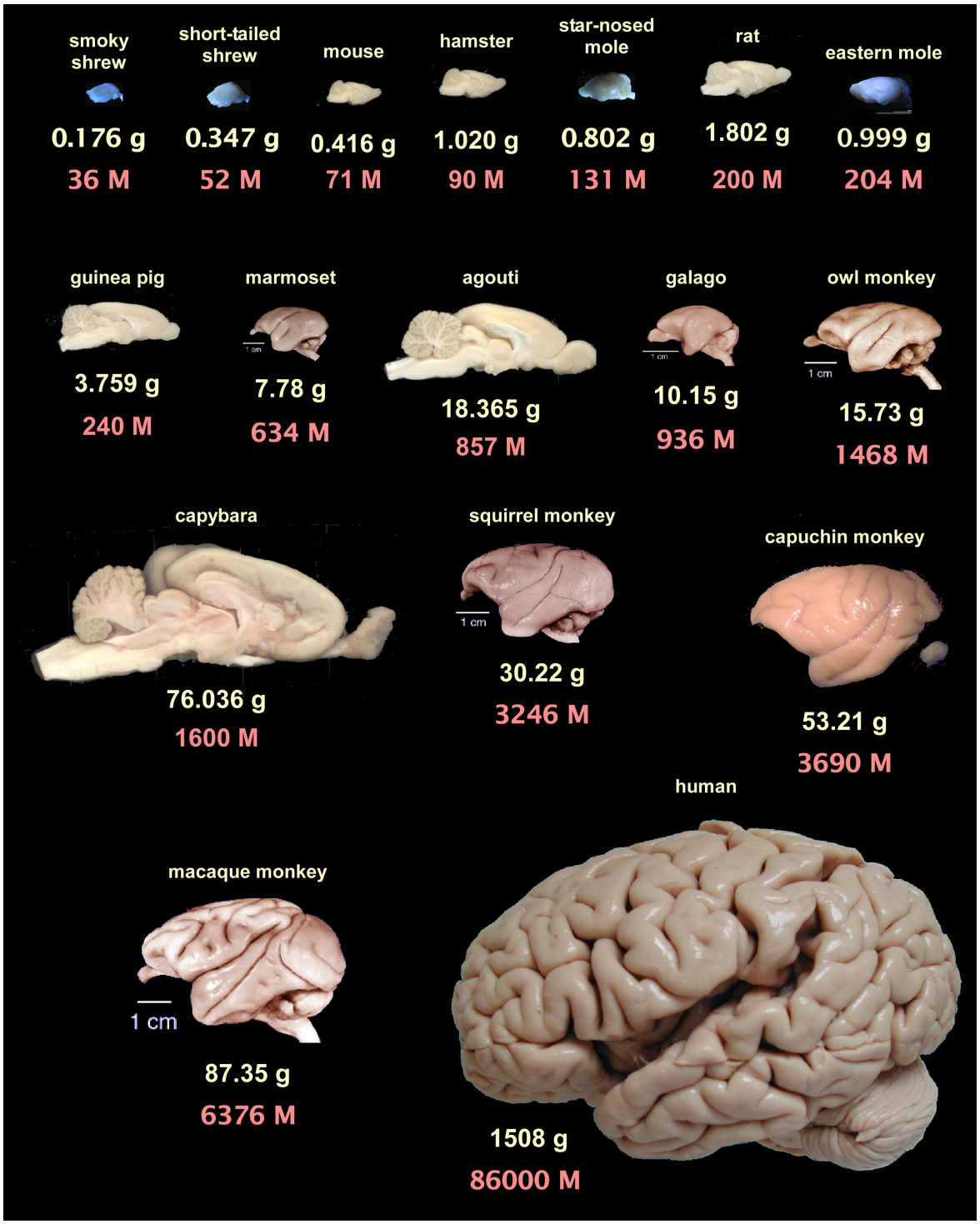


FIGURE 3 | Brain mass and total number of neurons for the mammalian species examined so far with the isotropic fractionator. Brains are arranged from left to right, top to bottom, in order of increasing number of neurons according to average species values from Herculano-Houzel et al., 2006 (rodents), Herculano-Houzel et al., 2007 (non-human primates), Sarko et al., 2009 (insectivores) and Azevedo et al., 2009 (human brain). Rodent brains face right, primate brains face left, insectivore brains can be identified in the figure by their bluish hue (due to illumination conditions). All images shown to the same scale. Primate images, except for the capuchin monkey and human brain, from the University of Wisconsin and Michigan State Comparative Mammalian Brain Collections (www.brainmuseum.org). Insectivore images kindly provided by Diana Sarko, and human brain image by Roberto Lent. Rodent images from the author. Notice that some rodent brains, such as the agouti and the capybara, contain fewer neurons than primate brains that are smaller than them.

as brain size changes, primate brains scale in size in a much more space-saving, economical manner compared to the inflationary growth that occurs in rodents, in which larger numbers of neurons are accompanied by larger neurons.

The cognitive consequences of this difference, which allows primate brains to enjoy the benefits of a large increase in numbers of neurons without the otherwise associated cost of a much larger increase in overall brain volume, can be glimpsed by returning to the comparison between rodents and primates of similar brain size. Now that absolute numbers of neurons can be compared across the similar-sized brains of agoutis and owl monkeys, and of capybaras and capuchin monkeys (**Figure 4**), the expected correlation between cognitive ability and numbers of neurons is actually found to hold: with 1468 million neurons, owl monkeys have almost twice as many neurons in the brain as agoutis (which hold 857 million), and about four times more neurons in the cerebral cortex than the agouti (442 million versus 113 million). Likewise, the capuchin monkey brain has more than twice the number of neurons of the larger-brained capybara (3.7 billion against 1.6 billion), and also about four times more neurons in the cerebral cortex (1.1 billion against 0.3).

The significance of the difference in scaling rules for building brains with large numbers of neurons becomes even more obvious if one considers the expected number of neurons for a generic rodent brain of human-sized proportions, weighing 1.5 kg: such a brain would have only 12 billion neurons, and a much larger number of 46 billion non-neuronal cells. This number of neurons is smaller than the number of neurons estimated to exist in the human cerebral cortex alone (Pakkenberg and Gundersen, 1997; Pelvig et al., 2008), and about seven times smaller than the number of neurons predicted for a 1.5-kg brain built with the scaling rules that apply to primates (see below).

Table 1 | Power law exponents that apply to the scaling of brain mass, or structure mass, as a function of the number of neurons they contain in rodents, insectivores and primates.

	Rodents	Insectivores	Primates
Brain mass × neurons	$M_{BR} \sim N_{BR}^{1.550}$	$M_{BR} \sim N_{BR}^{1.016}$	$M_{BR} \sim N_{BR}^{1.056}$
Cortical mass × neurons	$M_{CX} \sim N_{CX}^{1.744}$	$M_{CX} \sim N_{CX}^{1.520}$	$M_{CX} \sim N_{CX}^{1.077}$
Cerebellar mass × neurons	$M_{CB} \sim N_{CB}^{1.372}$	$M_{CB} \sim N_{CB}^{1.028}$	$M_{CB} \sim N_{CB}^{0.990}$
Remaining areas mass × neurons	$M_{RA} \sim N_{RA}^{1.153}$	$M_{RA} \sim N_{RA}^{0.926}$	$M_{RA} \sim N_{RA}^{1.013}$

Data are from Herculano-Houzel et al. (2006), Sarko et al. (2009) and Herculano-Houzel et al. (2009). Scaling laws for primate brains do not include human values.

THE CELLULAR COMPOSITION OF THE HUMAN BRAIN

The determination of the cellular scaling rules that apply to primate brains (Herculano-Houzel et al., 2007) enabled us to predict the cellular composition of the human brain. According to these rules, a generic primate brain of 1.5 kg should have 93 billion neurons, and 112 billion non-neuronal cells: glial cells, thus, should constitute at most half of all brain cells. This generic primate brain should have a cerebral cortex of about 1.4 kg, containing 25 billion neurons, and a cerebellum weighing 120 g, with 70 billion neurons (**Table 2**).

Establishing whether the human brain indeed conforms to the scaling rules that apply to other primates, however, required determining its cellular composition using the same method. This was accomplished by Azevedo et al. (2009), who found that the adult male human brain, at an average of 1.5 kg, has 86 billion neurons and 85 billion non-neuronal cells – numbers that deviate from the expected by 7 and 24% only. The human cerebral cortex, with an average 1233 g and 16 billion neurons, is slightly

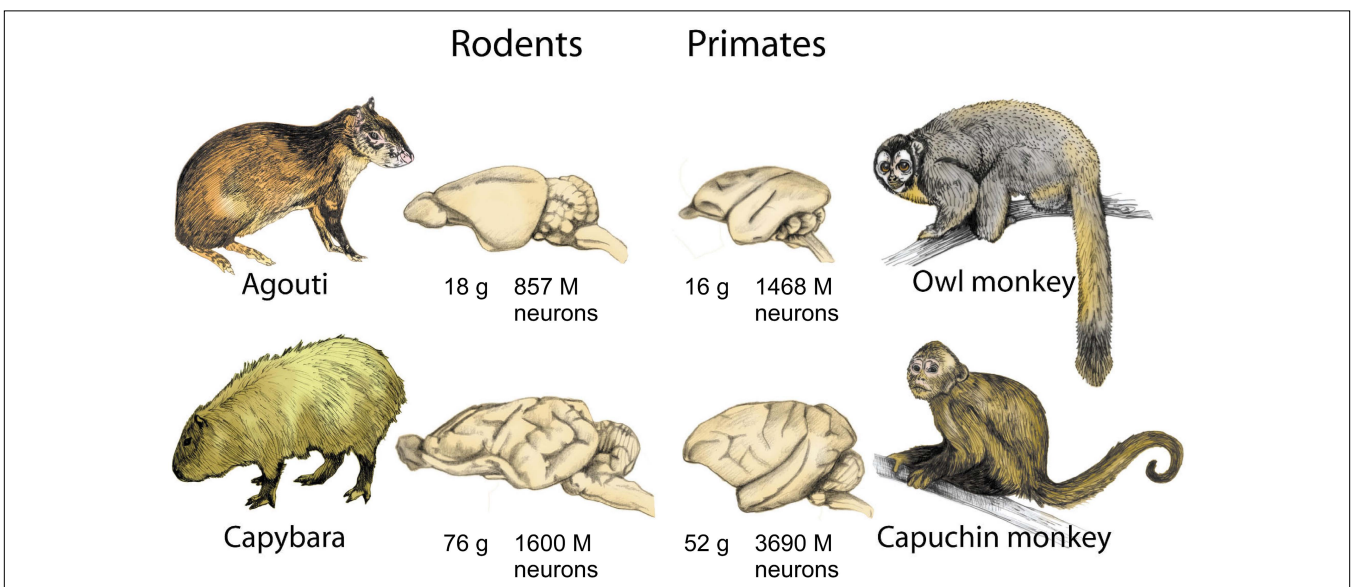


FIGURE 4 | Brain size is not a reliable indicator of number of neurons across orders. Because of the different cellular scaling rules that apply to rodent and primate brains, primates always concentrate larger numbers of neurons in the brain than rodents of a similar, or even larger, brain size. Data from Herculano-Houzel et al. (2006, 2007). Illustration by Lorena Kaz.

Table 2 | Expected values for a generic rodent and primate brains of 1.5 kg, and values observed for the human brain (Azevedo et al., 2009).

	Generic rodent brain	Generic primate brain	Human brain
Brain mass	1500 g	1500 g	1508 g
Total number of neurons in brain	12 billion	93 billion	86 billion
Total number of non-neurons in brain	46 billion	112 billion	85 billion
Mass, cerebral cortex	1154 g	1412 g	1233 g
Neurons, cerebral cortex	2 billion	25 billion	16 billion
Relative size of the cerebral cortex	77% of brain mass	94% of brain mass	82% of brain mass
Relative number of neurons in cerebral cortex	17% of brain neurons	27% of brain neurons	19% of brain neurons
Mass, cerebellum	133 g	121 g	154 g
Neurons, cerebellum	10 billion	61 billion	69 billion
Relative size of the cerebellum	9% of brain mass	8% of brain mass	10% of brain mass

Notice that although the expected mass of the cerebral cortex and cerebellum are similar for these hypothetical brains, the numbers of neurons that they contain are remarkably different. The human brain thus exhibits seven times more neurons than expected for a rodent brain of its size, but 92% of what would be expected of a hypothetical primate brain of the same size. Expected values were calculated based on the power laws relating structure size and number of neurons (irrespective of body size) that apply to average species values for rodents (Herculano-Houzel et al., 2006) and primate brains (Herculano-Houzel et al., 2007), excluding the olfactory bulb.

below expectations for a primate brain of 1.5 kg, while the human cerebellum, at 154 g and 69 billion neurons, matches or even slightly exceeds the expected (**Table 2**).

Although not observed in the comparatively small rodent species analyzed, the enlargement of the cerebral cortex is not, in principle, an exclusive feature of the human brain: a similar expansion of the mass of the cerebral cortex, relative to the whole brain, is predicted by both the rodent and primate cellular scaling rules, irrespective of the number of neurons contained in the cortex (**Table 2**). Remarkably, the human cerebral cortex, which represents 82% of brain mass, holds only 19% of all neurons in the human brain – a fraction that is similar to the fraction that we observed in several other primates, rodents, and even insectivores (**Figure 1**). The relatively large human cerebral cortex, therefore, is not different from the cerebral cortex of other animals in its relative number of neurons.

It should be noted that the unchanging proportional number of neurons in the cerebral cortex relative to the whole brain does not contradict an expansion in volume, function and number of neurons of the cerebral cortex in evolution: the absolute number of neurons in the rodent and primate cerebral cortex does increase much faster in larger brains compared to the number of neurons in the combined brainstem, diencephalon and basal ganglia, and is accompanied by a similarly fast increase in the number of neurons in the cerebellum (**Figure 5**).

Because of the diverging power laws that relate brain size and number of neurons across rodents and primates, the latter can hold more neurons in the same brain volume, with larger neuronal densities than found in rodents. Since neuronal density does not scale with brain size in primates, but decreases with increasing brain size in rodents, the larger the brain size, the larger is the difference in number of neurons across similar-sized rodent and primate brains.

PREDICTIONS FOR GREAT APES

The finding that the same cellular scaling rules apply to humans and non-anthropoid primate brains alike, irrespective of body size, indicates that the brains of the great apes, which diverged from the hominin lineage before humans, should also conform to the same cellular scaling rules. An examination of the cellular composition of the cerebellum of orangutans and one gorilla shows that the sizes of the cerebellum and cerebral cortex predicted for these species from the number of cells in the cerebellum match their actual sizes, which suggests that the brain of these animals indeed is built according to the same scaling rules that apply to humans and other primates (Herculano-Houzel and Kaas, in preparation). In view of the discrepant relationship between body and brain size in humans, great apes, and non-anthropoid primates, these findings suggest that the rules that apply to scaling primate brains are much more conserved than those that apply to scaling the body. This raises the possibility that brain mass and body mass across species are only correlated, rather than brain mass being determined by body mass, as presumed in studies that focus on the variation of residuals after regression onto body size. Supportive evidence comes from the dissociation between brain and body growth in development, in which the former actually precedes the latter (reviewed in Deacon, 1997), and from our observation that body mass seems more free to vary across species than brain mass as a function of its number of neurons. In this view, it will be interesting to consider the alternative hypothesis that body size is not a determining variable for brain size in comparative studies of brain neuroanatomy, and particularly not an (independent) parameter for assessing quantitative aspects of the human brain.

DO WE HAVE THE MOST NEURONS? PREDICTIONS FOR OTHER LARGE-BRAINED MAMMALS

The different cellular scaling rules that apply to rodents and primates strongly indicate that it is not valid to use brain size as a proxy for number of neurons across humans, whales, elephants and other large-brained species belonging to different mammalian orders. One consequence of this realization is that sheer size alone, or in relation to body size, is not an adequate parameter to qualify, or disqualify, the human brain as "special".

A comparison of expected numbers can nevertheless be very illuminating. For instance, given the cellular scaling rules that we have observed for rodents (Herculano-Houzel et al., 2006), a hypothetical

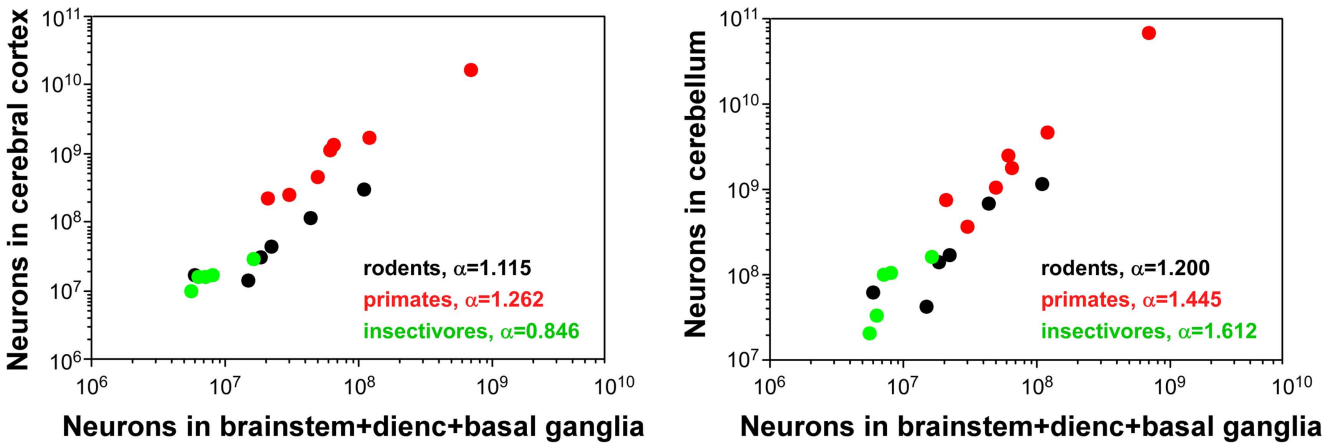


FIGURE 5 | Numbers of neurons increase faster in the cerebral cortex and cerebellum than in the remaining brain areas (the combined brainstem, diencephalon and basal ganglia). Data points indicate average values for individual species of rodents (Herculano-Houzel et al., 2006), primates (Herculano-Houzel et al., 2007), including humans (Azevedo et al., 2009), and insectivores (Sarko et al., 2009).

rodent brain with 86 billion neurons, like the human brain, would be predicted to weigh overwhelming 35 kg – a value that is way beyond the largest known brain mass of 9 kg for the blue whale, and probably physiologically unattainable. As mentioned above, a generic rodent brain of human-sized proportions, weighing 1.5 kg, would have only 12 billion neurons: in this sense, therefore, being a primate endows us with seven times more neurons than would be expected if we were rodents. Notice that this remarkable difference does not rely on assumptions about how brain size or cellular composition relate to body size in the species.

A burning question is now whether cetaceans and elephants, endowed with much larger brains than humans, also have much larger numbers of neurons than humans. According to one estimate, the false killer whale and the African elephant would have about 11 billion neurons in the cerebral cortex, despite their large size – and fewer neurons than the 11.5 billion estimated by the same method for the human cerebral cortex, though only marginally so (Roth and Dicke, 2005). These estimates, however, were obtained by simply multiplying cerebral cortical volume and the neuronal densities determined for a few cortical areas, which probably do not reflect average neuronal density in the entire cortex.

Although direct measurements of cellular composition are not yet available from whole elephant and whale brains, it is illuminating to consider how their cellular compositions would differ depending on whether predicted from the scaling rules that apply to rodent or to primate brains. As shown in **Table 3**, the difference in numbers of neurons predicted to compose the brains of the false killer whale and of the African elephant is 10-fold depending on the scaling rules employed. Speculatively, the estimate of neuronal density in the gray matter of the cerebral cortex of the whale and the elephant at a low figure of about 7000 neurons/mm^3 (Tower, 1954) suggests that these brains conform to scaling rules that are much closer to those that apply to rodents than to the primate scaling rules. It may turn out, therefore, these very large brains are composed of remarkably fewer neurons than the human brain, despite their size, thanks to the distinct, economical scaling rules that apply to primates in general (and not to humans in particular).

THE HUMAN BRAIN IS A LINEARLY SCALED-UP PRIMATE BRAIN IN ITS NUMBER OF NEURONS. WHAT NOW?

COGNITIVE ABILITIES, BRAIN SIZE AND NUMBER OF NEURONS

To conclude that the human brain is a linearly scaled-up primate brain, with just the expected number of neurons for a primate brain of its size, is not to state that it is unremarkable in its capabilities. However, as studies on the cognitive abilities of non-human primates and other large-brained animals progress, it becomes increasingly likely that humans do not have truly unique cognitive abilities, and hence must differ from these animals not qualitatively, but rather in the combination and extent of abilities such as theory of mind, imitation and social cognition (Marino et al., 2009). Quantitative changes in the neuronal composition of the brain could therefore be a main driving force that, through the exponential combination of processing units, and therefore of computational abilities, leads to events that may look like "jumps" in the evolution of brains and intelligence (Roth and Dicke, 2005). Such quantitative changes are likely to be warranted by increases in the absolute (rather than relative) numbers of neurons in relevant cortical areas and, coordinately, in the cerebellar circuits that interact with them (Ramnani, 2006). Moreover, viewing the human brain as a linearly scaled-up primate brain in its cellular composition does not diminish the role that particular neuroanatomical arrangements, such as changes in the relative size of functional cortical areas (for instance, Semendeferi et al., 2001; Rilling and Seligman, 2002), in the volume of prefrontal white matter (Schoenemann et al., 2005) or in the size of specific portions of the cerebellum (Ramnani, 2006) may play in human cognition. Rather, such arrangements should contribute to brain function in combination with the large number of neurons in the human brain. Our analysis of numbers of neurons has so far been restricted to large brain divisions, such as the entire cerebral cortex and the ensemble of brainstem, diencephalon and basal ganglia, but an analysis of the cellular scaling of separate functional cortical areas and the related subcortical structures is underway. Such data should allow us to address important issues such as mosaic evolution through concerted changes in the functionally related

Table 3 | Predicted cellular composition of whale and elephant brains if they scaled according to rodent or primate cellular scaling rules.

	Predicted from rodent rules	Predicted from primate rules
FALSE KILLER WHALE, 3650 G		
Neurons, whole brain	21 billion	212 billion
Mass, cerebral cortex	3000 g	3655 g
Neurons, cerebral cortex	3 billion	55 billion
Neuronal density, cortex	1000 neurons/mg*	30–80,000 neurons/mg**
Mass, cerebellum	304 g	279 g
Neurons, cerebellum	19 billion	140 billion
Neuronal density, cerebellum	63,500 neurons/mg*	400–600,000 neurons/mg*
AFRICAN ELEPHANT, 4200 G		
Neurons, whole brain	23 billion	241 billion
Mass, cerebral cortex	3488 g	4245 g
Neurons, cerebral cortex	3 billion	62 billion
Neuronal density, cortex	960 neurons/mg*	30–80,000 neurons/mg***
Mass, cerebellum	347 g	318 g
Neurons, cerebellum	21 billion	159 billion
Neuronal density, cerebellum	61,200 neurons/mg*	400–600,000 neurons/mg*

Neuronal densities (*) predicted from the rodent scaling rules apply to the whole structures, including white matter. Neuronal densities predicted from the primate scaling rules are the range observed in primate gray matter (**) (Herculano-Houzel et al., 2008), and in the primate cerebellum including white matter (***), since neuronal density does not covary with structure size in primates (Herculano-Houzel et al., 2007). Notice the difference in predicted numbers of neurons depending on the scaling rules applied. Given the low neuronal densities observed in the whale and elephant gray matter, of about 7000 neurons/mm^3 of gray matter, it is reasonable to speculate that the scaling rules that apply to whale and elephant brains are closer to the rules that apply to rodent brains than to the rules that apply to primate brains. Notice also that the actual size of the elephant cerebellum, at about 1 kg (Hakeem et al., 2005), is much larger than the predicted here.

components of distributed systems, and the presumed increase in relative number of neurons in systems that increase in importance (Barton and Harvey, 2000; Barton, 2006).

If cognitive abilities among non-human primates scale with absolute brain size (Deaner et al., 2007) and brain size scales linearly across primates with its number of neurons (Herculano-Houzel et al., 2007), it is tempting to infer that the cognitive abilities of a primate, and of other mammals for that matter, are directly related to the number of neurons in its brain. In this sense, it is interesting to realize that, if the same linear scaling rules are considered to apply to great apes as to other primates, then similar three-fold differences in brain size and in brain neurons alike apply to humans compared to gorillas, and to gorillas compared to baboons. This, however, is not to say that any cognitive advantages that the human brain may have over the gorilla and that the gorilla may have over the baboon are equally three-fold – although these differences are difficult to quantify. Since neurons interact combinatorially through the synapses they establish with one another, and further so as they interact in networks, the increase in cognitive abilities afforded by increasing the number of neurons in the brain can be expected to increase exponentially with absolute number of neurons, and might even be subject to a thresholding effect once critical points of information processing are reached. In this way, the effects of a three-fold increase in numbers of neurons may be much more remarkable when comparing already large brains, such as those of humans and gorillas, than when comparing small brains, such as those of squirrel monkeys and galagos.

INTRASPECIFIC VARIABILITY IN SIZE, NUMBERS AND ABILITIES

One final caveat to keep in mind when studying scaling of numbers of brain neurons, particularly in regard to cognition, is that relationships observed across species need not apply to comparisons across individuals of the same species. Not only the extent of intraspecific variation is much smaller (on the order of 10–50%) than interspecific variation (which spans five orders of magnitude within mammals; Tower, 1954; Stolzenburg et al., 1989), but also the mechanisms underlying interspecific and intraspecific variation are also likely to differ. Our own preliminary data suggest that, indeed, variations in brain size across rats of the same age are not correlated with variations in numbers of neurons (Morterá and Herculano-Houzel, unpublished observations). There is no justification, therefore, to extend the linear correlation between brain size and number of neurons across primates to a putative correlation across persons of different brain sizes (which might be used, inappropriately, as grounds for claims that larger-brained individuals have more neurons, and are therefore "smarter", than smaller-brained persons). In fact, although men have been reported to have more neurons in the cerebral cortex than women (Pakkenberg and Gundersen, 1997; Pelvig et al., 2008), there is no significant correlation between brain size and general cognitive ability within families (Schoenemann et al., 2000). Across these individuals, other factors such as variations in number and identity of synaptic connections within and across structures, building on a statistically normal, albeit variable, number of neurons, and depending on genetics and life experiences such as learning, are more likely to be determinant of the individual cognitive abilities (see, for instance, Mollgaard et al., 1971; Black et al., 1990; Irwin et al., 2000; Draganski et al., 2004).

CONCLUDING REMARKS: OUR PLACE IN NATURE

Novel quantitative data on the cellular composition of the human brain and its comparison to other primate brains strongly indicate that we need to rethink our notions about the place that the human brain holds in nature and evolution, and rewrite some of the basic concepts

that are taught in textbooks. Accumulating evidence (Deacon, 1997; Roth and Dicke, 2005; Deaner et al., 2007) indicates that an alternative view of the source of variations in cognitive abilities across species merits investigation: one that disregards body and brain size and examines absolute numbers of neurons as a more relevant parameter instead. Now that these numbers can be determined in various brains and their structures, direct comparisons can be made across species and orders, with no assumptions about body–brain size relationships required. Complementarily, however, it now becomes possible to examine how numbers of neurons in the brain, rather than brain size, relate to body mass and surface as well as metabolism, parameters that have been considered relevant in comparative studies (Martin, 1981; Fox and Wilczynski, 1986; MacLarnon, 1996; Schoenemann, 2004), in order to establish what mechanisms underlie the loosely correlated scaling of body and brain.

According to this now possible neuron-centered view, rather than to the body-centered view that dominates the literature (see Gazzaniga, 2008, for a comprehensive review), the human brain has the number of neurons that is expected of a primate brain of its size; a cerebral cortex that is exactly as large as expected for a primate brain of 1.5 kg; just as many neurons as expected in the cerebral cortex for the size of this structure; and, despite having a relatively large cerebral cortex (which, however, a rodent brain of 1.5 kg would also be predicted to have), this enlarged cortex holds just the same proportion of brain neurons in humans as do other primate cortices (and rodent cortices, for that matter). This final observation calls for a reappraisal of the view of brain evolution that concentrates on the expansion of the cerebral cortex, and its replacement with a more integrated view of coordinate evolution of cellular composition, neuroanatomical structure, and function of cerebral cortex and cerebellum (Whiting and Barton, 2003).

Other "facts" that deserve updating are the ubiquitous quote of 100 billion neurons (a value that lies outside of the margin of variation found so far in human brains; Azevedo et al., 2009), and, more strikingly, the widespread remark that there are 10× more glial cells than neurons in the human brain. As we have shown, glial cells in the human brain are at most 50% of all brain cells, which is an important finding since it is one more brain characteristic that we share with other primates (Azevedo et al., 2009).

Finally, if being considered the bearer of a linearly scaled-up primate brain does not sound worthy enough for the animal that considers himself the most cognitively able on Earth, one can note that there are, indeed, two advantages to the human brain when compared to others – even if it is not an outlier, nor unique in any remarkable way. First, the human brain scales as a primate brain: this economical property of scaling alone, compared to rodents, assures that the human brain has many more neurons than would fit into a rodent brain of similar size, and possibly into any other similar-sized brain. And second, our standing among primates as the proud owners of the largest living brain assures that, at least among primates, we enjoy the largest number of neurons from which to derive cognition and behavior as a whole. It will now be interesting to determine whether humans, indeed, have the largest number of neurons in the brain among mammals as a whole.

ACKNOWLEDGMENTS

Thanks to Roberto Lent and Jon Kaas for continued support and encouragement, to Bruno Mota for insightful discussions, and to the colleagues who participated in the quantification of the cellular composition of the brain of various species. Supported by CNPq-Ed. Universal and FAPERJ-Jovem Cientista do Estado do Rio de Janeiro.

REFERENCES

Allen, N. J., and Barres, B. A. (2009). Glia – more than just brain glue. *Nature* 457, 675–677.

Andersen, B. B., Korbo, L., and Pakkenberg, B. (1992). A quantitative study of the human cerebellum with unbiased stereological techniques. *J. Comp. Neurol.* 326, 549–560.

Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., Jacob Filho, W., Lent, R., and Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* 513, 532–541.

Barton, R. A. (2002). How did brains evolve? *Nature* 415, 134–135.

Barton, R. A. (2006). Primate brain evolution: integrating comparative, neurophysiological, and ethological data. *Evol. Anthropol.* 15, 224–236.

Barton, R. A., and Harvey, P. H. (2000). Mosaic evolution of brain structure in mammals. *Nature* 405, 1055–1058.

Black, J. E., Isaacs, K. R., Anderson, B. J., Alcantara, A. A., and Greenough, W T. (1990). Learning causes synaptogenesis, whereas motor activity causes angiogenesis, in cerebellar cortex of adult rats. *Proc. Natl. Acad. Sci. USA* 87, 5568–5572.

Butti, C., Sherwood, C. C., Hakeem, A. Y., Allman, J. M., and Hof, P. R. (2009). Total number and volume of Von Economo neurons in the cerebral cortex of cetaceans. *J. Comp. Neurol.* 515, 243–259.

Clark, D. A., Mitra, P. P., and Wang, S. S. (2001). Scalable architecture in mammalian brains. *Nature* 411, 189–193.

Deacon, T. W. (1997). What makes the human brain different? *Annu. Rev. Anthropol.* 26, 337–357.

Deaner, R. O., Isler, K., Burkart, J., and van Schaik, C. (2007). Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. *Brain Behav. Evol.* 70, 115–124.

Doetsch, F. (2003). The glial identity of neural stem cells. *Nat. Neurosci.* 6, 1127–1134.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., and May, A. (2004). Changes in grey matter induced by training. *Nature* 427, 311–312.

Finlay, B. L., and Darlington, R. B. (1995). Linked regularities in the development and evolution of mammalian brains. *Science* 268, 1578–1584.

Fox, J. H., and Wilczynski, W. (1986). Allometry of major CNS divisions: towards a reevaluation of somatic brain-body scaling. *Brain Behav. Evol.* 28, 157–169.

Friede, R. (1954). Quantitative share of the glia in development of the cortex. *Acta Anat.* 20, 290–296.

Gazzaniga, M. (2008). Human: The Science Behind What Makes us Unique. New York, Harper Collins.

Hakeem, A. Y., Hof, P. R., Sherwood, C. C., Switzer R. C. III, Rasmussen, L. E. L., and Allman, J. M. (2005). Brain of the African elephant (*Loxodonta africana*): neuroanatomy from magnetic resonance images. *Anat. Rec.* 287A, 1117–1127.

Hakeem, A. Y., Sherwood, C. C., Bonar, C. J., Butti, C., Hof, P. R., and Allman, J. M. (2009). Von Economo neurons in the elephant brain. *Anat. Rec.* 292, 242–248.

Harvey, P. H., and Pagel, M. D. (1991). Modeling physiological and anthropometric variables known to vary with body size and other confounding variables. *Yearb. Phys. Anthropol.* 48, 141–153.

Haug, H. (1987). Brain sizes, surfaces, and neuronal sizes of the cortex cerebri: a stereological investigation of man and his variability and a comparison with some mammals (primates, whales, marsupials, insectivores, and one elephant). *Am. J. Anat.* 180, 126–142.

Hawking, A., and Olszewski, J. (1957). Glia/nerve cell index for cortex of the whale. *Science* 126, 76–77.

Herculano-Houzel, S. (2002). Do you know your brain? A survey on public neuroscience literacy at the closing of the decade of the brain. *Neuroscientist* 8, 98–110.

Herculano-Houzel, S. (2007). Encephalization, neuronal excess, and neuronal index in rodents. *Anat. Rec.* 290, 1280–1287.

Herculano-Houzel, S., Collins, C. E., Wong, P., and Kaas, J. H. (2007). Cellular scaling rules for primate

brains. *Proc. Natl. Acad. Sci. USA* 104, 3562–3567.

Herculano-Houzel, S., Collins, C. E., Wong, P., Kaas, J. H., and Lent, R. (2008). The basic nonuniformity of the cerebral cortex. *Proc. Natl. Acad. Sci. USA* 105, 12593–12598.

Herculano-Houzel, S., and Lent, R. (2005). Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain. *J. Neurosci.* 25, 2518–2521.

Herculano-Houzel, S., Mota, B., and Lent, R. (2006). Cellular scaling rules for rodent brains. *Proc. Natl. Acad. Sci. USA* 103, 12138–12143.

Irwin, S. A., Galvez, R., and Greenough, W. T. (2000). Dendritic spine structural anomalies in fragile-X mental retardation syndrome. *Cereb. Cortex* 10, 1038–1044.

Jerison, H. (1973). Evolution of the Brain and Intelligence. New York, Academic Press.

Jerison, H. (1985). Animal intelligence as encephalization. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 308, 21–35.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). Principles of Neural Science, 4th Edn. New York, McGraw-Hill, pp. 19–20.

Lange, W. (1975). Cell number and cell density in the cerebellar cortex of man and some other mammals. *Cell Tissue Res.* 157, 115–125.

Lefebvre, L., Reader, S. M., and Sol, D. (2004). Brains, innovations and evolution in birds and primates. *Brain Behav. Evol.* 63, 233–246.

MacDonald, D. W. (1981). Dwindling resources and the social behavior of capybaras (*Hydrochaeris hydrochaeris*) (Mammalia). *J. Zool. Lond.* 194, 371–391.

MacLarnon, A. (1996). The scaling of gross dimensions of the spinal cord in primates and other species. *J. Hum. Evol.* 30, 71–87.

Marino, L. (1998). A comparison of encephalization between odontocete cetaceans and anthropoid primates. *Brain Behav. Evol.* 51, 230–238.

Marino, L. (2002). Convergence of complex cognitive abilities in cetaceans and primates. *Brain Behav. Evol.* 59, 21–32.

Marino, L., Connor, R. C., Fordyce, R. E., Herman, L. M., Hof, P. R., Lefebvre, L., Lusseau, D., McCowan, B., Nimchinsky, E. A., Pack, A. A., Rendell, L., Reidenberg, J. S., Reiss, D., Uhen, M. D., Van der Gucht, E., and Whitehead, H. (2009). Cetaceans have complex brains for complex cognition. *PLoS Biol.* 5:e139. doi:10.1371/journal.pbio.0050139

Martin, R. D. (1981). Relative brain size and basal metabolic rate in terrestrial vertebrates. *Nature* 293, 57–60.

Mollgaard, K., Diamond, M. C., Bennett, E. L., Rosenzweig, M. R., and Lindner, B. (1971). Quantitative synaptic changes with differential experience in rat brain. *Int. J. Neurosci.* 2, 113–127.

Nimchinsky, E. A., Glissen, E., Allman, J. M., Perl, D. P., Erwin, J. M., and Hof, P. R. (1999). A neuronal morphologic type unique to humans and great apes. *Proc. Natl. Acad. Sci. USA* 96, 5268–5273.

Nishiyama, A., Yang, Z., and Butt, A. (2005). What's in a name? *J. Anat.* 207, 687–693.

Noctor, S. C., Martinez-Cerdeno, V., and Kriegstein, A. R. (2007). Contribution of intermediate progenitor cells to cortical histogenesis. *Arch. Neurol.* 64, 639–642.

Nunn, C., and Barton, R. A. (2000). Allometric slopes and independent contrasts: a comparative test of Kleiber's law in primates. *Am. Nat.* 156, 519–533.

Pakkenberg, B., and Gundersen, H. J. G. (1988). Total numbers of neurons and glial cells in human brain nuclei estimated by the disector and the fractionator. *J. Microsc.* 150, 1–20.

Pakkenberg, B., and Gundersen, H. J. G. (1997). Neocortical neuron number in humans: effect of sex and age. *J. Comp. Neurol.* 384, 312–320.

Pelvig, D. P., Pakkenberg, H., Stark, A K., and Pakkenberg, B. (2008). Neocortical glial cell numbers in human brains. *Neurobiol. Aging* 29, 1754–1762.

Ramnani, N. (2006). The primate cortico-cerebellar system: anatomy and function. *Nat. Neurosci.* 7, 511–522.

Reichenbach, A. (1989). Glia:neuron index: review and hypothesis to account for different values in various mammals. *Glia* 2, 71–77.

Rilling, J. K. (2006). Human and nonhuman primate brains: are they allometrically scaled versions of the same design? *Evol. Anthropol.* 15, 65–77.

Rilling, J. K., and Seligman, R. A. (2002). A quantitative morphometric comparative analysis of the primate temporal lobe. *J. Hum. Evol.* 42, 505–533.

Roth, G., and Dicke, U. (2005). Evolution of the brain and intelligence. *Trends Cogn. Sci.* 9, 250–257.

Sarko, D. K., Catania, K. C., Leitch, D. B., Kaas, J. H., and Herculano-Houzel, S. (2009). Cellular scaling rules of insectivore brains. *Front. Neuroanat.* 3, 8. doi:10.3389/neuro.05.008.2009

Schoenemann, P. T. (2004). Brain size scaling and body composition in mammals. *Brain Behav. Evol.* 63, 47–60.

Schoenemann, P. T., Budinger, T. F., Sarich, V. M., and Wang, W. S. (2000). Brain size does not predict general cognitive ability within families. *Proc. Natl. Acad. Sci. USA* 97, 4932–4937.

Schoenemann, P. T., Sheehan, M. J., and Glotzer, L. D. (2005). Prefrontal white matter volume is disproportionately larger in humans than in other primates. *Nat. Neurosci.* 8, 242–252.

Semendeferi, K., Armostrong, E., Schleicher, A., Zilles, K., and Van Hoesen, G. W. (2001). Prefrontal cortex in humans and apes: a comparative study of area 10. *Am. J. Phys. Anthropol.* 114, 224.

Semendeferi, K., Lu, A., Schenker, N., and Damasio, H. (2002). Humans and great apes share a large frontal cortex. *Nat. Neurosci.* 5, 272–276.

Shariff, G. A. (1953). Cell counts in the primate cerebral cortex. *J. Comp. Neurol.* 98, 381–400.

Sherwood, C. C., Stimpson, C. D., Raghanti, M. A., Wildman, D. E., Uddin, M., Grossman, L. I., Goodman, M., Redmond, J. C., Bonar, C. J., Erwin, J. M., and Hof, P. R. (2006). Evolution of increased glia-neuron ratios in the human frontal cortex. *Proc. Natl. Acad. Sci. USA* 103, 13606–13611.

Stolzenburg, J. U., Reichenbach, A., and Neumann, M. (1989). Size and density of glial and neuronal cells within the cerebral neocortex of various insectivorian species. *Glia* 2, 78–84.

Sultan, F. (2002). Analysis of mammalian brain architecture. *Nature* 415, 133–134.

Tower, D. B. (1954). Structural and functional organization of mammalian cerebral cortex: the correlation of neurone density with brain size; cortical neurone density in the fin whale (*Balaenoptera physalus* L.) with a note on the cortical neurone density in the Indian elephant. *J. Comp. Neurol.* 101, 19–51.

Tower, D. B., and Elliot, K. A. (1952). Activity of acetylcholine system in cerebral cortex of various unanesthetized mammals. *Am. J. Physiol.* 168, 747–759.

Ullian, E. M., Sapperstein, S. K., Christopherson, K. S., and Barres, B. A. (2001). Control of synapse number by glia. *Science* 291, 657–660.

Vallender, E. J. (2008). Exploring the origins of the human brain through molecular evolution. *Brain Behav. Evol.* 72, 168–177.

Whiting, B. A., and Barton, R. A. (2003). The evolution of the cortico-cerebellar complex in primates: anatomical connections predict patterns of correlated evolution. *J. Hum. Evol.* 44, 3–10.

Williams, R. W., and Herrup, K. (1988). The control of neuron number. *Annu. Rev. Neurosci.* 11, 423–453.

Zhang, K., and Sejnowski, T. J. (2000). A universal scaling law between gray matter and white matter of cerebral cortex. *Proc. Natl. Acad. Sci. USA* 97, 5621–5626.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 July 2009; paper pending published: 05 August 2009; accepted: 29 September 2009; published online: 09 November 2009.

Citation: Herculano-Houzel S (2009) The human brain in numbers: a linearly scaled-up primate brain. Front. Hum. Neurosci. ***3****:31. doi: 10.3389/neuro.09.031.2009*

Copyright © 2009 Herculano-Houzel. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

The role of inflammation in depression: from evolutionary imperative to modern treatment target

Andrew H. Miller[1] and Charles L. Raison[2]

Abstract | Crosstalk between inflammatory pathways and neurocircuits in the brain can lead to behavioural responses, such as avoidance and alarm, that are likely to have provided early humans with an evolutionary advantage in their interactions with pathogens and predators. However, in modern times, such interactions between inflammation and the brain appear to drive the development of depression and may contribute to non-responsiveness to current antidepressant therapies. Recent data have elucidated the mechanisms by which the innate and adaptive immune systems interact with neurotransmitters and neurocircuits to influence the risk for depression. Here, we detail our current understanding of these pathways and discuss the therapeutic potential of targeting the immune system to treat depression.

Conspecifics
Members of the same species.

Depression is a devastating disorder, afflicting up to 10% of the adult population in the United States and representing one of the leading causes of disability worldwide[1]. Although effective treatments are available, approximately one third of all patients with depression fail to respond to conventional antidepressant therapies[2], further contributing to the global burden of the disease. Accordingly, there is a pressing need for new conceptual frameworks for understanding the development of depression to develop better treatments. In this Review, we outline emerging data that point to the immune system — and, in particular, the inflammatory response — as a potentially important contributor to the pathophysiology of depression. We first consider the origins of this notion from an evolutionary perspective, examining the advantages of depressive behaviours in the context of host immune responses to pathogens, predators and conspecifics in ancestral environments. The pivotal role of psychosocial stress in the modern world are then examined, highlighting inflammasome activation and immune cell trafficking as novel mechanisms by which stress-induced inflammatory signals can be transmitted to the brain. Neurotransmitters and neurocircuits that are targets of the inflammatory response are also explored followed by an examination of brain–immune interactions as risk and resilience factors for depression. Finally, these interactions are discussed as a foundation for a new era of therapeutics that target the immune system to treat depression, with a focus on how immunological biomarkers can be used to personalize care.

An evolutionary perspective

Data from humans and laboratory animals provide compelling evidence that stress-relevant neurocircuitry and immunity form an integrated system that evolved to protect organisms from a wide range of environmental threats. For example, in the context of a laboratory stressor that entails delivering a speech to a judgmental panel of supposed 'behavioural experts', subjects experience a classic 'fight or flight' response characterized by increases in heart rate and blood pressure as well as in cortisol and catecholamines. But something else happens within the body that demands a deeper explanation. The stressor activates key inflammatory pathways in peripheral blood mononuclear cells, including activation of the transcription factor nuclear factor-κB (NF-κB), and leads to marked increases in circulating levels of pro-inflammatory cytokines, such as interleukin-6 (IL-6)[3,4]. In essence, the body mounts an immune response not against a pathogen, but against a threat to the subject's self-esteem. Moreover, individuals at high risk of developing depression (for example, those who have experienced early-life trauma) show increased inflammatory responses to such laboratory stressors compared with low-risk individuals[3]. Furthermore, the greater the inflammatory response to a psychosocial stressor, the more probable the subject is to develop depression over the ensuing months[5]. Two questions immediately present themselves: why should a stimulus devoid of any pathogen induce an inflammatory response, and why should this response promote the development of depression?

[1]*Emory University School of Medicine, Winship Cancer Institute, Atlanta, 30322 Georgia, USA.*
[2]*School of Human Ecology, University of Wisconsin–Madison, Madison, 53706 Wisconsin, USA.*

Correspondence to A.H.M.
amill02@emory.edu

doi:10.1038/nri.2015.5
Published online 29 Dec 2015

© 2015 Macmillan Publishers Limited. All rights reserved

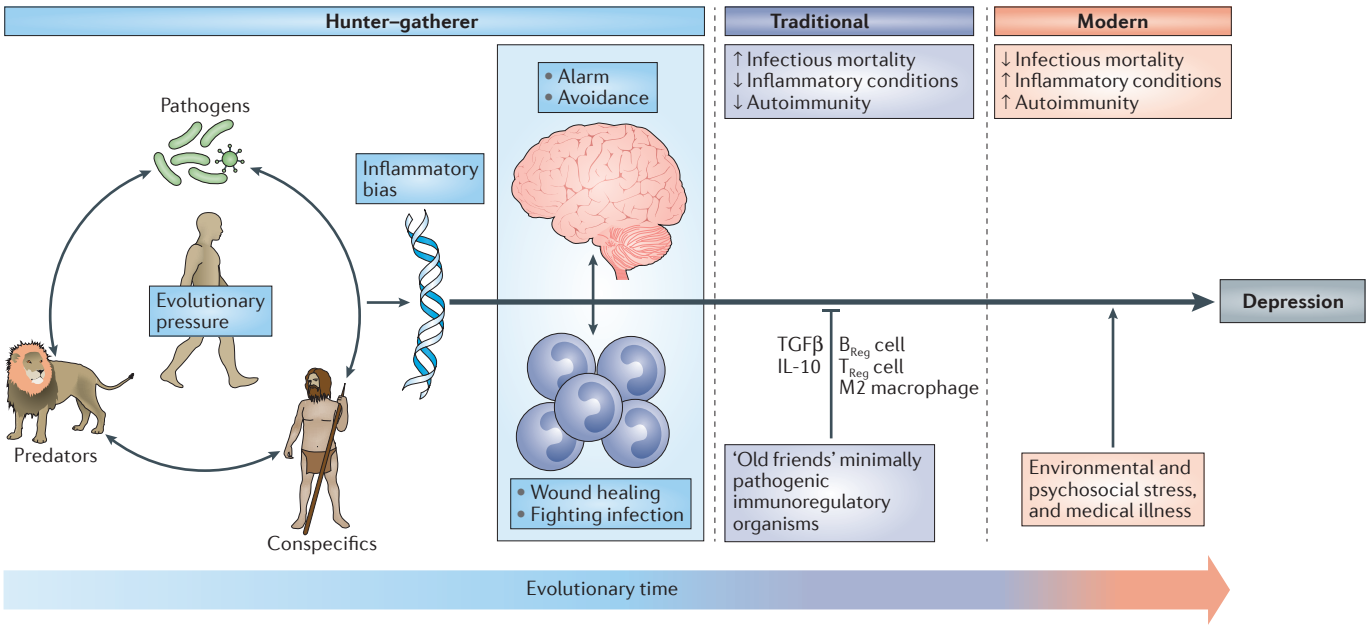


Figure 1 | **Evolutionary legacy of an inflammatory bias.** Early evolutionary pressures derived from human interactions with pathogens, predators and human conspecifics (such as rivals) resulted in an inflammatory bias that included an integrated suite of immunological and behavioural responses that conserved energy for fighting infection and healing wounds, while maintaining vigilance against attack. This inflammatory bias is believed to have been held in check during much of human evolution by exposure to minimally pathogenic, tolerogenic organisms in traditional (that is, rural) environments that engendered immunological responses characterized by the induction of regulatory T (T_{Reg}) cells, regulatory B (B_{Reg}) cells and immunoregulatory M2 macrophages as well as the production of the anti-inflammatory cytokines interleukin-10 (IL-10) and transforming growth factor-β (TGFβ). In modern times, sanitized urban environments of more developed societies are rife with psychological challenges but generally lacking in the types of infectious challenges that were primary sources of morbidity and mortality across most of human evolution. In the absence of traditional immunological checks and balances, the psychological challenges of the modern world instigate ancestral immunological and behavioural repertoires that represent a decided liability, such as high rates of various inflammation-related disorders including depression.

Pathogen host defence and depression. No coherent answer to these questions is apparent if immunity is viewed as merely another physiological system within the body. However, when seen against the back-drop of millions of years of co-evolution between mammals and the world of microorganisms and parasites, the human inflammatory bias exposed by laboratory stressors and reflected in the association between immune activation and depression not only makes imminent adaptive sense but also provides insight into a paradox deep within the heart of depression itself; namely, why are the genetic alleles that are most frequently associated with depression so common in the modern gene pool[6] (FIG. 1)?

Most adaptive theories of depression have focused on the potential benefits of depressive symptoms for relationships with other humans[7]. However, recent models have shifted the focus away from relationships with people, to relationships — both detrimental and beneficial — with pathogens[6,8]. These theories, which are supported by converging evidence (BOX 1), posit that modern humans have inherited a genomic bias towards inflammation because this response — and the depressive symptoms it promotes — enhanced host survival and reproduction in the highly pathogenic environments in which humans evolved[6]. From this theoretical perspective, at least some of the human vulnerability to depression evolved out of a behavioural repertoire — often referred to as 'sickness behaviour' — which promoted host survival in the face of infection. Indeed, it has been hypothesized that the social avoidance and anhedonia characteristic of depression serve to shunt energy resources to fighting infection and wound healing, whereas the hypervigilance characteristic of anxiety disorders, commonly co-morbid with depression, subserves protection from attack and subsequent pathogen exposure[6,9]. Even psychological stress can be understood from this theoretical perspective, given that the vast majority of stressors faced by mammals over evolutionary time boiled down to risks inherent in hunting, being hunted or competing for reproductive access or status. In all of these circumstances, the risk of pathogen invasion — and subsequent death from infection — was greatly increased as a result of wounding. In ancestral environments, the association between stress perception and risk of subsequent wounding was reliable enough that evolution favoured organisms that prepotently activated inflammatory systems in response to a wide array of environmental threats and challenges (including psychosocial stressors), even if this activation was often a 'false alarm' (REF. 6).

Sickness behaviour
An adaptive response to illness, often precipitated by infection, that includes social withdrawal, decreased appetite, lethargy, impaired concentration, depressed mood, irritability, muscle aches and pain, and fever. This syndrome is believed to prioritize shifting of energy resources to fighting infection and wound healing.

Anhedonia
A lack of interest in usually pleasurable activities that represents a decrease in motivation, which can either represent a decrease in the response to reward or in the willingness to expend effort to obtain reward.

© 2015 Macmillan Publishers Limited. All rights reserved

Box 1 | **Pathogen host defence hypothesis of depression**

Several lines of evidence support the notion that the evolution and persistence of depression risk alleles and depressive symptoms in human populations are based on their relevance to 'pathogen host defence'. This evidence includes:

- Until recently, approximately 50% of humans died from infectious causes before adulthood, thereby providing strong selective pressure for genetic alleles that enhance host defence[124].
- As a result of strong selective pressure, microbial interactions have been a primary driver of human evolution[125].
- Patterns of inflammatory activation associated with depression promote survival in highly pathogenic environments while increasing mortality in sanitary conditions common in the developed world[126].
- The best replicated risk alleles for depression have pro-inflammatory and/or anti-pathogen protective effects or have been implicated in social behaviours that are likely to reduce pathogen exposure[6].
- Environmental risk factors for the development of depression (that is, psychosocial stress, early life adversity, obesity and processed-food diet) are uniformly pro-inflammatory[13].
- Exposure to pro-inflammatory cytokines produces a sickness syndrome with symptoms that overlap considerably with those seen in depression and that can be ameliorated by treatment with antidepressants[23]. In addition, the onset of depression is often mistaken with development of sickness, and symptoms associated with infections are often mistaken with the onset of depression[127].
- Chronic cytokine exposure produces a combination of withdrawal and/or energy conservation, anxiety and/or hypervigilance behaviours and emotions that commonly coexist in depression[6,9].
- Symptoms shared by depression and sickness behaviour — such as hyperthermia and reduced iron availability — that lack any conceivable social value have potent anti-pathogen effects[6].

The 'pathogen host defence' hypothesis of depression may also provide insight into the twofold increase in depression in women compared to men, especially during the reproductive years[10]. Recent data indicate that women are more sensitive to the behavioural effects of inflammation, demonstrating greater increases in depressed mood than men following endotoxin exposure despite a similar magnitude in cytokine (IL-6 and tumour necrosis factor (TNF)) responses[11]. Women also exhibit a greater likelihood than men to develop depression in response to standardized doses of interferon-α (IFNα)[12]. By being more sensitive to inflammation-induced depressive symptoms, women may have benefited more from the protection provided by these symptoms in terms of fighting infection, healing wounds and avoiding subsequent pathogen exposure. Given the potentially negative impact of inflammation on reproductive success (for example, by reducing fertility and impairing lactation), the increase of depressive symptoms in women across evolutionary time may have given women of reproductive age an advantage in coping with and avoiding pathogens and the related inflammation, with increased depressive disorders being the ultimate trade-off in modern times.

Modern exaggeration of the inflammatory bias. The prevalence of autoimmune, allergic and inflammatory diseases has markedly increased in the past 100 years, and rates of these conditions follow a similar upward trajectory in societies transitioning from traditional (that is, rural) to modern (that is, urban) ways of life[13]. Increasing evidence suggests that this pattern of widespread immune dysregulation may result from disruptions in our relationship and/or contact with a variety of co-evolved, non-lethal immunoregulatory microorganisms and parasites, especially commensals and symbiotes in the microbiotas of the gut, skin and nasal and oral cavities, that were ubiquitous in the natural environments in which humans evolved[14]. Although widely disparate, these organisms (often referred to as 'old friends') share a tendency to reduce inflammation and suppress effector immune cells through the induction of IL-10 and transforming growth factor-β (TGFβ) while promoting the development of anti-inflammatory immune cell populations, such as alternatively activated (also referred to as 'M2') macrophages and regulatory T (T_{Reg}) cells and regulatory B cells[13,14] (FIG. 1). Owing to various cultural changes, including the loss of exposure to microbial diversity with the advent of sanitation practices, modern humans now lack this immunoregulatory input — especially during infancy and childhood. Consequently, we find ourselves in a condition of an exacerbated inflammatory bias, with the particular conditions afflicting any given individual largely the result of genetic predisposition and environmental (for example, psychosocial) exposures[13,14], ultimately accounting for the high co-morbidity between depression and autoimmune, allergic and inflammatory disorders[13,15].

Inflammation and depression

Data supporting the role of inflammation in depression are extensive and include findings that span experimental paradigms. Patients with major depressive disorder exhibit all of the cardinal features of an inflammatory response, including increased expression of pro-inflammatory cytokines and their receptors and increased levels of acute-phase reactants, chemokines and soluble adhesion molecules in peripheral blood and cerebrospinal fluid (CSF)[16,17]. Peripheral blood gene expression profiles consistent with a pro-inflammatory 'M1' macrophage phenotype and an over-representation of IL-6, IL-8 and type I IFN-induced signalling pathways have also been described[18–20]. In addition, increased expression of a variety of innate immune genes and proteins, including IL-1β, IL-6, TNF, Toll-like receptor 3 (TLR3) and TLR4, has been found in post-mortem brain samples from suicide victims that had depression[16,18,19,21]. Meta-analyses of the literature conclude that peripheral blood IL-1β, IL-6, TNF and C-reactive protein (CRP) are the most reliable biomarkers of inflammation in patients with depression[16]. Polymorphisms in inflammatory cytokine genes, including those encoding IL-1β, TNF and CRP, have also been associated with depression and its response to treatment[22]. Moreover, other genes implicated in depression derived from meta-analyses of genome-wide association studies have been linked to the immune response and the response to pathogens including TNF[6] (BOX 1). Administration of inflammatory cytokines (for example, IFNα) or their inducers (for example, endotoxin or typhoid vaccination) to otherwise non-depressed individuals causes

Major depressive disorder
A clinical syndrome of depression characterized by the primary symptoms of depressed mood and anhedonia, and diagnosed using criteria set forth by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

© 2015 Macmillan Publishers Limited. All rights reserved

symptoms of depression[23–26]. Furthermore, blockade of cytokines, such as TNF, or of inflammatory signalling pathway components, such as cyclooxygenase 2, has been shown to reduce depressive symptoms in patients with medical illnesses, including rheumatoid arthritis, psoriasis and cancer, as well as in patients with major depressive disorder[27–29].

As the field has matured, it has become increasingly apparent that inflammatory markers are elevated not only in a subgroup of patients with depression[30,31] but also in patients with other neuropsychiatric disorders including anxiety disorders and schizophrenia[32,33]. Moreover, as described below, it may be more accurate to characterize the impact of inflammation on behaviour as being associated not wholly with depression but with specific symptom dimensions across diagnoses that align with the Research Domain Criteria framework put forth by the National Institute of Mental Health (US Department of Health and Human Services). These symptoms, including positive and negative valence systems, relate to altered motivation and motor activity (anhedonia, fatigue and psychomotor impairment) and increased threat sensitivity (anxiety, arousal and alarm)[34]. Finally, inflammation has been associated with antidepressant treatment non-responsiveness[9,32,35–37]. For example, in a recent study, 45% of patients with non-response to conventional antidepressants exhibited a CRP >3 mg L^{-1} (REF. 30), which is considered indicative of a high level of inflammation on the basis of widely accepted cut-off points[38]. Of note, however, the percentage of patients with high CRP levels can vary as a function of the population being studied, with higher percentages in patients with depression and treatment resistance, childhood maltreatment, medical illnesses and metabolic syndrome.

Immune pathways involved in depression

Inflammasomes: stress in translation. Exposure to psychosocial stress is one of the most robust and reproducible predictors of developing depression in humans and is the primary experimental pathway to depressive-like behaviour in laboratory animals. Thus, the observation that exposure to a psychosocial laboratory stressor can activate an inflammatory response in humans was a major breakthrough in linking inflammation to depression[3,4]. An important question for the field, however, is by what mechanism is stress translated into inflammation? Although considerable attention has been paid to stress-induced neuroendocrine pathways, including the hypothalamic–pituitary–adrenal (HPA) axis and the sympathetic nervous system (SNS), both of which have immunomodulatory functions[39], recent focus has been shifted towards inflammasomes, which may represent a crucial immunological interface between stress and inflammation[40] (FIG. 2). Inflammasomes are cytosolic protein complexes that form in myeloid cells in response to pathogenic microorganisms and non-pathogenic or 'sterile' stressors. Assembly of the inflammasome leads to activation of caspase 1, which then cleaves the precursor forms of IL-1β and IL-18 into the active cytokines[41]. Given the relatively sterile nature of psychosocial stress, primary interest has been directed towards understanding how inflammasome activation in depression may be triggered by endogenous damage-associated molecular patterns (DAMPs), including ATP, heat shock proteins (HSPs), uric acid, high mobility group box 1 (HMGB1) and a variety of molecules linked with oxidative stress. Indeed, all of these DAMPs are induced by the psychological and mixed (that is, psychological and physiological) stressors used in animal models of depression[42]; an effect that is in part mediated by stress-induced release of catecholamines[43]. Moreover, studies in laboratory animals indicate that chronic mild-stress activates the NOD-, LRR- and pyrin domain-containing protein 3 (NLRP3) inflammasome, which is well-known to respond to DAMPs[44,45]. Blockade of NLRP3 reverses stress-induced increases in IL-1β in the peripheral blood and brain, while also abrogating depressive-like behaviour in mice[45]. Interestingly, NLRP3 inflammasome upregulation and caspase-mediated cleavage of the glucocorticoid receptor can cause resistance to the effects of glucocorticoids, which are among the most potent anti-inflammatory hormones in the body[46,47]. Stress-induced glucocorticoid resistance is a well-characterized biological abnormality in patients with major depressive disorder and has been associated with increased inflammation[48,49].

Supporting the potential role of the NLRP3 inflammasome in human depression are data demonstrating that increased expression of NLRP3 and caspase 1 in peripheral blood mononuclear cells of patients with depression is associated with increased blood concentrations of IL-1β and IL-18, which in turn correlate with depression severity[19,50]. In addition, DAMPs that are known to activate NLRP3 are increased in patients with mood disorders, with examples including HSPs, reactive oxygen species and other markers of oxidative stress such as xanthine oxidase, peroxides and F2-isoprostanes[51–53]. Finally, there is increasing interest in the potential role of the gut microbiome in mood regulation, which may be mediated in part by the inflammasomes[54]. Indeed, non-pathogenic commensal bacteria and derived microbial-associated molecular patterns (MAMPs) in the gut can leak into the peripheral circulation during stress and activate the inflammasomes[55], a process mediated by the SNS and catecholamines[56] (FIG. 2). Of note, stress-induced increases in IL-1β and IL-18 were attenuated by treating animals with antibiotics or neutralizing lipopolysaccharide (LPS), demonstrating the importance of the composition of the gut microbiome and gut permeability in stress-induced inflammatory responses[55]. Taken together, these data support the notion that the inflammasome may be a key immunological point of integration of stress-induced danger signals that ultimately drive inflammatory responses relevant to depression.

Transmitting inflammatory signals to the brain. In addition to increased expression of innate immune cytokines and TLRs in post-mortem brain samples from suicide victims with depression, evidence of microglial and astroglial activation in several brain regions including frontal cortex, anterior cingulate cortex (ACC)

© 2015 Macmillan Publishers Limited. All rights reserved

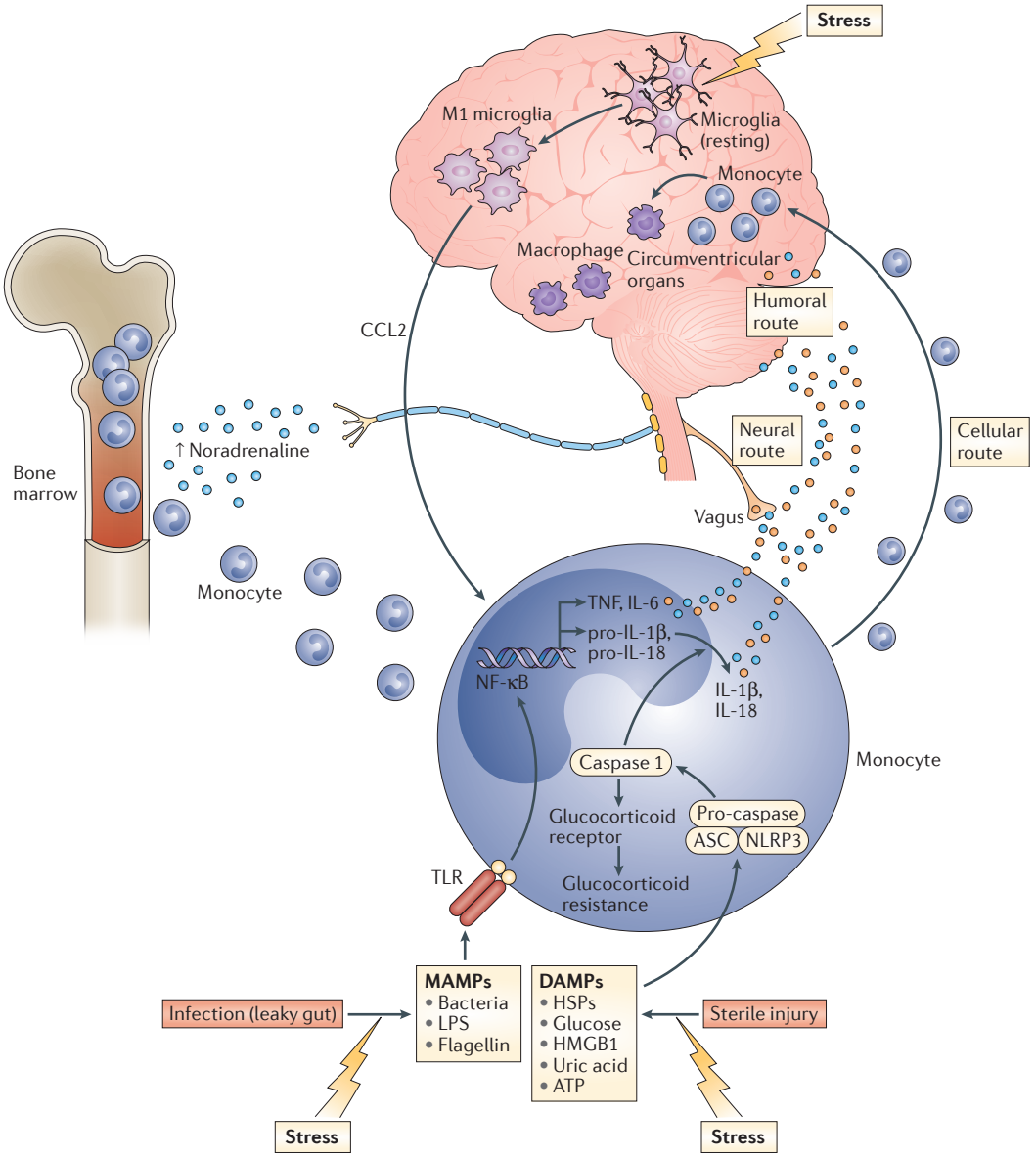


Figure 2 | **Transmitting stress-induced inflammatory signals to the brain.** In the context of psychosocial stress, catecholamines (such as noradrenaline) released by activated sympathetic nervous system fibres stimulate bone marrow production and the release of myeloid cells (for example, monocytes) that enter the periphery where they encounter stress-induced damage-associated molecular patterns (DAMPs), bacteria and bacterial products such as microbial-associated molecular patterns (MAMPs) leaked from the gut. These DAMPs and MAMPs subsequently activate inflammatory signalling pathways such as nuclear factor-κB (NF-κB) and the NOD-, LRR- and pyrin domain-containing protein 3 (NLRP3) inflammasome. Stimulation of NLRP3 in turn activates caspase 1, which leads to the production of mature interleukin-1β (IL-1β) and IL-18 while also cleaving the glucocorticoid receptor contributing to glucocorticoid resistance. Activation of NF-κB stimulates the release of other pro-inflammatory cytokines including tumour necrosis factor (TNF) and IL-6, which together with IL-1β and IL-18 can access the brain through humoral and neural routes. Psychosocial stress can also lead to the activation of microglia to a M1 pro-inflammatory phenotype, which release CC-chemokine ligand 2 (CCL2) that in turn attracts activated myeloid cells to the brain via a cellular route. Once in the brain, activated macrophages can perpetuate central inflammatory responses. ASC, apoptosis-associated speck-like protein containing a CARD; HMGB1, high mobility group box 1; HSP, heat shock protein; LPS, lipopolysaccharide; TLR, Toll-like receptor.

and thalamus in post-mortem studies of patients with depression have been described[57–59,60]. Moreover, a well-controlled neuroimaging study using positron emission tomography (PET) and a radiolabelled tracer for the translocator protein (TSPO) — which is overexpressed in activated microglia, macrophages and astrocytes — revealed increased immune activation in the brains of patients with major depressive disorder compared with control subjects[61]. Of note, not all studies have revealed increased TSPO binding in patients with depression, possibly owing to effects of medication and/or a paucity of subjects with increased inflammation[61,62]. However,

© 2015 Macmillan Publishers Limited. All rights reserved

data from endotoxin administration to healthy volunteers indicates that radiolabelled TSPO ligands can readily identify cellular activation in several regions of the brain following a potent peripheral immune stimulus[63].

Work from laboratory animal studies has elucidated several pathways through which inflammatory signals can be transmitted from the periphery to the brain (FIG. 2). These data support the idea that inflammatory responses in peripheral tissues may drive inflammation in the brain leading to depression. Much of the early work focused on how inflammatory cytokines, which are relatively large molecules, could cross the blood–brain barrier (BBB) and influence brain function[64]. Two major pathways have been described: the 'humoral pathway', which involves cytokine passage through leaky regions in the BBB, such as the circumventricular organs, and the binding of cytokines to saturable transport molecules on the BBB; and the 'neural pathway', which involves the binding of cytokines to peripheral afferent nerve fibres, such as the vagus nerve, that in turn stimulate ascending catecholaminergic fibres in the brain and/or are translated back into central cytokine signals[16]. More recently, however, attention has shifted to a third pathway referred to as the 'cellular pathway', which involves the trafficking of activated immune cells, typically monocytes, to the brain vasculature and parenchyma. The details of this pathway have been elegantly dissected in the context of behavioural changes in mice that are associated with peripherally induced inflammation in the liver[65]. In these studies, the release of TNF from inflamed liver was found to stimulate microglial cell production of CC-chemokine ligand 2 (CCL2; also known as MCP1) that then attracted monocytes to the brain[65]. Blockade of monocyte infiltration to the brain using antibodies specific for the adhesion molecules P-selectin and α4 integrin abrogated depressive-like behaviour in this animal model[65]. Of note, cytokine-stimulated astrocytes also may be major producers of chemokines, such as CCL2 and CXC-chemokine ligand 1 (CXCL1), that attract immune cells to the brain[66]. The cellular pathway additionally has been elucidated in the context of social defeat stress, whereby GFP-labelled monocytes coalesced in several regions of the brain associated with the detection of threat (for example, amygdala) — an effect that was dependent on CCL2 and was facilitated by mobilization of monocytes from the bone marrow as a result of stress-induced release of catecholamines[67,68] (FIG. 2). Of note, initial microglial activation during social defeat stress appeared to be a result of neuronal activation by catecholamines and decreased neuronal production of CX_3C-chemokine ligand 1 (CX_3CL1; also known as fractalkine), which maintains microglia in a quiescent state[67,68]. Interestingly, this cellular pathway has received intriguing support from post-mortem analyses of brain tissue from patients with depression who committed suicide that showed increased numbers of perivascular macrophages in association with increased gene expression of allograft inflammatory factor 1 (AIF1, also known as IBA1) and CCL2, which are associated with macrophage activation and cellular trafficking[59].

Social defeat stress
A model of depression that entails repeated exposure to a conspecific animal screened for aggressive behaviour. The animals are placed together in the same cage where they are exposed to brief bouts of defeat lasting 5–10 minutes daily typically for 6–10 days.

This evidence of peripheral myeloid cells trafficking to the brain during depression constitutes some of the first data supporting the existence of a central inflammatory response in human depression that is primarily driven by peripheral inflammatory events. Moreover, data demonstrate that antibodies that are specific for TNF but which do not cross the BBB, can block stress-induced depression in mice[69]. These findings indicate that peripheral inflammatory responses not only can provide important clues to the immunological mechanisms of inflammation in depression but also may serve as biomarkers and targets of immune-based therapies for depression. Protein biomarkers such as plasma CRP and TNF as well as immunotherapies targeting individual cytokines such as TNF, IL-1 and IL-6 may be most relevant in this regard. Of note, plasma CRP is a strong response predictor in anti-cytokine therapy[70].

Cytokines and neurotransmitters. Given the pivotal importance of neurotransmission to mood regulation, attention has been paid to the impact of inflammation and inflammatory cytokines on the monoamines serotonin, noradrenaline and dopamine, as well as on the excitatory amino acid glutamate (FIG. 3). There are several pathways through which inflammatory cytokines can lead to reduced synaptic availability of the monoamines, which is believed to be a fundamental mechanism in the pathophysiology of depression[71]. For example, IL-1β and TNF induction of p38 mitogen-activated protein kinase (MAPK) has been shown to increase the expression and function of the reuptake pumps for serotonin, leading to decreased synaptic availability of serotonin and depressive-like behaviour in laboratory animals[72]. Through the generation of reactive oxygen and nitrogen species, inflammatory cytokines have also been found to decrease the availability of tetrahydrobiopterin (BH4), a key enzyme co-factor in the synthesis of all monoamines that is highly sensitive to oxidative stress[73]. Indeed, CSF concentrations of BH4 have been shown to be negatively correlated with CSF levels of IL-6 in patients treated with the inflammatory cytokine IFNα[74]. In addition, the plasma phenylalanine to tyrosine ratio, an indirect measure of BH4 activity, was shown to correlate with CSF concentrations of dopamine as well as symptoms of depression in IFNα-treated patients[74]. Activation of the enzyme indoleamine 2,3-dioxygenase (IDO) is also believed to be involved in cytokine-induced neurotransmitter alterations, in part by diverting the metabolism of tryptophan (the primary amino acid precursor of serotonin) into kynurenine, a compound that can be converted into the neurotoxic metabolite quinolinic acid by activated microglia and infiltrating monocytes and macrophages in the brain[75,76]. Of note, increased levels of quinolinic acid have been found in microglia in the ACC of suicide victims who suffered from depression[77]. Quinolinic acid directly activates receptors for glutamate (that is, *N*-methyl-D-aspartate (NMDA) receptors) while also stimulating glutamate release and blocking glutamate reuptake by astrocytes[78]. The effects of quinolinic acid on glutamate converge with the direct effects of

© 2015 Macmillan Publishers Limited. All rights reserved

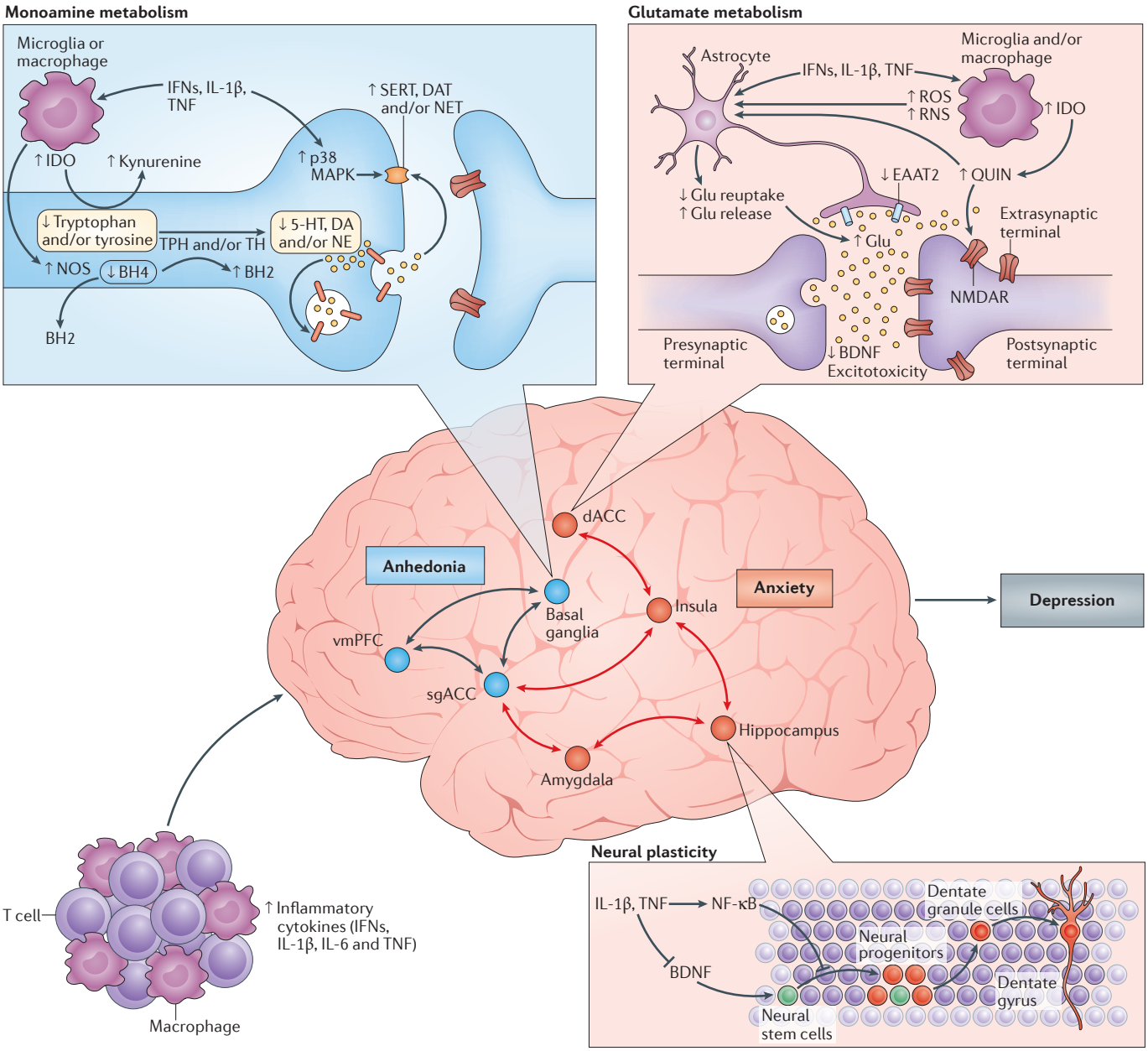


Figure 3 | **Cytokine targets in the brain: neurotransmitters and neurocircuits.** Once in the brain, the inflammatory response can affect metabolic and molecular pathways influencing neurotransmitter systems that can ultimately affect neurocircuits that regulate behaviour, especially behaviours relevant to decreased motivation (anhedonia), avoidance and alarm (anxiety), which characterize several neuropsychiatric disorders including depression. On a molecular level, pro-inflammatory cytokines including type I and II interferons (IFNs), interleukin-1β (IL-1β) and tumour necrosis factor (TNF) can reduce the availability of monoamines — serotonin (5-HT), dopamine (DA) and noradrenaline (NE) — by increasing the expression and function of the presynaptic reuptake pumps (transporters) for 5-HT, DA and NE through activation of mitogen-activated protein kinase (MAPK) pathways and by reducing monoamine synthesis through decreasing enzymatic co-factors such as tetrahydrobiopterin (BH4), which is highly sensitive to cytokine-induced oxidative stress and is involved in the production of nitric oxide (NO) by NO synthase (NOS). Many cytokines, including IFNγ, IL-1β and TNF, can also decrease relevant monoamine precursors by activating the enzyme indoleamine 2,3-dioxygenase (IDO), which breaks down tryptophan, the primary precursor for serotonin, into kynurenine. Activated microglia can convert kynurenine into quinolinic acid (QUIN), which binds to the *N*-methyl-D-aspartate receptor (NMDAR), a glutamate (Glu) receptor, and together with cytokine-induced reduction in astrocytic Glu reuptake and stimulation of astrocyte Glu release, in part by induction of reactive oxygen species (ROS) and reactive nitrogen species (RNS), can lead to excessive Glu, an excitatory amino acid neurotransmitter. Excessive Glu, especially when binding to extrasynaptic NMDARs, can in turn lead to decreased brain-derived neurotrophic factor (BDNF) and excitotoxicity. Inflammation effects on growth factors such as BDNF in the dentate gyrus of the hippocampus can also affect fundamental aspects of neuronal integrity including neurogenesis, long-term potentiation and dendritic sprouting, ultimately affecting learning and memory. Cytokine effects on neurotransmitter systems, especially DA, can inhibit several aspects of reward motivation and anhedonia in corticostriatal circuits involving the basal ganglia, ventromedial prefrontal cortex (vmPFC) and subgenual and dorsal anterior cingulate cortex (sgACC and dACC, respectively), while also activating circuits regulating anxiety, arousal, alarm and fear including the amygdala, hippocampus, dACC and insula. BH2, dihydrobiopterin; DAT, dopamine transporter; EAAT2, excitatory amino acid transporter 2; NET, noradrenaline transporter; NF-κB, nuclear factor-κB; SERT, serotonin transporter; TH, tyrosine hydroxylase; TPH, tryptophan hydroxylase. Copyrighted 2015. Advanstar. 120580:1115BN.

© 2015 Macmillan Publishers Limited. All rights reserved

pro-inflammatory cytokines on glutamate metabolism that include decreasing the expression of astrocyte glutamate reuptake pumps and stimulating astrocytic glutamate release[79], ultimately contributing to excessive glutamate both within and outside the synapse. The binding of glutamate to extrasynaptic NMDA receptors leads to increased excitotoxicity and decreased production of brain-derived neurotrophic factor (BDNF)[80]. BDNF fosters neurogenesis, an important prerequisite for an antidepressant response, and has been shown to be reduced by IL-1β and TNF and their downstream signalling pathways including NF-κB in stress-induced animal models of depression[81,82]. Increased levels of glutamate in the basal ganglia and dorsal ACC (dACC) — as measured by magnetic resonance spectroscopy (MRS) — have been described in patients receiving IFNα, and higher levels of glutamate correlated with an increase in depressive symptoms[83]. More recent data indicate that in patients with depression, increased inflammation as reflected by a CRP >3 mg L^{-1} is also associated with increased basal ganglia glutamate (compared with patients with a CRP <1 mg L^{-1}) that correlated with anhedonia and decreased psychomotor speed[84]. Interestingly, blocking glutamate receptors with ketamine or inhibiting IDO activity protects mice from LPS- or stress-induced depressive-like behaviour but leaves the inflammatory response intact[85,86]. These results indicate that increased activation of glutamate receptors by glutamate and/or quinolinic acid may be a common pathway through which inflammation causes depressive-like behaviour, suggesting that drugs that block glutamate receptor signalling and/or activation of the IDO pathway and its downstream metabolites might have unique applicability to patients with depression and increased inflammation. Importantly, conventional antidepressant medications act by increasing synaptic availability of monoamines and increasing neurogenesis through induction of BDNF[87]. Therefore, cytokines such as IL-1β and TNF serve to undermine these activities as they decrease the synaptic availability of monoamines while also decreasing BDNF and increasing extracellular glutamate, which is not a target of conventional antidepressant therapy. These cytokine-driven effects may explain the observations that increased inflammation is associated with less robust antidepressant treatment responses and that treatment-resistant patients exhibit increased inflammatory markers[88].

Effects of inflammation on neurocircuitry. Given the impact of cytokines on neurotransmitter systems that regulate the functional activity of neurocircuits throughout the brain, it is no surprise that neuroimaging studies have revealed cytokine-induced alterations in regional brain activity. Consistent with the evolutionary advantages of the partnership between the brain and the immune system, primary cytokine targets in the CNS involve those brain regions that regulate motivation and motor activity (promoting social avoidance and energy conservation) as well as arousal, anxiety and alarm (promoting hypervigilance and protection against attack) (FIG. 3).

Dopamine has a fundamental role in motivation and motor activity, and cytokines have been shown to decrease the release of dopamine in the basal ganglia in association with decreased effort-based motivation as well as reduced activation of reward circuitry in the basal ganglia, in particular the ventral striatum[89–91]. Inflammatory stimuli have been associated with reductions in reward responsiveness in the striatum across many neuroimaging platforms, demonstrating the validity and reproducibility of these cytokine-mediated effects on the brain in otherwise non-depressed individuals peripherally administered IFNα, endotoxin or typhoid vaccination and imaged by PET, functional magnetic resonance imaging (fMRI), MRS and quantitative magnetization transfer imaging[83,89,90,92,93]. Interestingly, recent fMRI studies suggest that inflammation-induced decreases in responsiveness to positive reward are also associated with increased sensitivity to aversive stimuli (that is, negative reinforcement) and reduced responsiveness to novelty in the substantia nigra (which is another dopamine-rich structure in the basal ganglia)[93,94]. Typhoid vaccination has also been shown to activate the subgenual ACC (sgACC), a brain region implicated in depression, and to decrease connectivity of the sgACC with the ventral striatum, an effect modulated by plasma IL-6 (REF. 26). These fMRI findings have recently been extended to patients with depression whose increased plasma CRP level is associated with decreased functional connectivity within reward-related circuits including the ventral striatum and the ventromedial prefrontal cortex that, in turn, mediates the relationship between CRP and anhedonia[95]. Indeed, patients with depression with a CRP >3 mg L^{-1} had little, if any, connectivity within reward-related circuits as measured by fMRI, whereas connectivity in patients with depression with a CRP <1 mg L^{-1} was similar to healthy controls[95]. Taken together, these data support the notion that the effect of cytokines on the brain in general and dopaminergic pathways in particular lead to a state of decreased motivation or anhedonia, which is a core symptom of depression.

fMRI studies have demonstrated that increased inflammation is also associated with increased activation of threat- and anxiety-related neurocircuitry, including the dACC as well as the insula and amygdala[26,96,97]. Of note, the dACC and amygdala are regions that exhibit increased activity in patients with high-trait anxiety and neuroticism[98], conditions that often accompany depression and are associated with increased inflammation. For example, increased concentrations of oral IL-6 and soluble TNF receptor 2 (also known as TNFRSF1B) in response to a public speaking stressor was significantly correlated with the response of the dACC to a social rejection task[97]. In addition, increased oral IL-6 expression in response to a social evaluation stressor was significantly correlated with activation of the amygdala, with subjects who exhibited the highest IL-6 responses to stress demonstrating the greatest connectivity within threat circuitry, including the amygdala and the dorsomedial prefrontal cortex, as measured by fMRI[99]. Interestingly, these data are consistent with the trafficking of monocytes to the amygdala during social defeat stress in mice[68].

© 2015 Macmillan Publishers Limited. All rights reserved

Risk and resilience

Increased inflammation and the risk for depression. Consistent with the emerging recognition that inflammation may cause depression in certain subgroups of individuals, epidemiological studies on large community samples — as well as smaller samples of medically ill individuals — have demonstrated that increased inflammation serves as a risk factor for the future development of depression. For example, increased peripheral blood CRP and IL-6 concentrations were found to significantly predict depressive symptoms after 12 years of follow up in the Whitehall II study of over 3,000 individuals, whereas no association was found between the presence of depressive symptoms and subsequent blood CRP and IL-6 levels[100]. Similar findings were reported in the English Longitudinal Study of Ageing in which a CRP >3 mg L^{-1} predicted depressive symptoms and not vice versa[101]. Of note, however, some studies have found no longitudinal relationship between depression and inflammation, and others have found that depression leads to increased inflammation[102]. Other factors known to be associated with increased peripheral inflammation, including childhood and adult trauma, have also been shown to be predictive of a greater risk of developing depression[103,104].

Both genetic and epigenetic mechanisms may explain why childhood or adult traumas can contribute to exaggerated or persistent inflammation and, ultimately, depression. For example, polymorphisms in *CRP* were associated not only with increased peripheral blood concentrations of CRP but also with symptoms of post-traumatic stress disorder, especially heightened arousal, in individuals exposed to civilian trauma[32]. Moreover, gene–environment interactions have been found to influence depression severity in response to chronic interpersonal stress: individuals carrying polymorphisms in *IL1B* that are associated with higher expression of peripheral IL-1β exhibited more severe depressive symptoms in the context of interpersonal stress than individuals without the *IL1B* risk allele[105]. Similarly, mice in which peripheral blood leukocytes produced high concentrations of LPS-induced IL-6 *ex vivo* before stress exposure showed decreased social exploration after social defeat stress, whereas mice that produced low levels of IL-6 before stress exposure exhibited no behavioural effects in response to social defeat[88]. Of note, adoptive transfer of bone marrow progenitor cells from mice producing high levels of IL-6 *ex vivo* to mice that produced low levels of IL-6 made these formerly stress-resilient animals sensitive to the depressive effects of social defeat[88].

Epigenetic changes in genes related to inflammation may also affect the risk for depression and anxiety in the context of psychosocial stress. Indeed, the well-documented association of childhood trauma with increased inflammation is linked to stress-induced epigenetic changes in *FKBP5*, a gene implicated in the development of depression and anxiety as well as in the sensitivity to glucocorticoids[106]. Allele-specific, childhood trauma-dependent DNA demethylation in functional glucocorticoid response elements of *FKBP5* were found to be associated with decreased sensitivity of peripheral blood immune cells to the inhibitory effects of the synthetic glucocorticoid dexamethasone on LPS-induced production of IL-6 *in vitro*[106]. Of note, decreased activation of glucocorticoid receptor-responsive genes in association with increased activation of genes regulated by NF-κB has been found to be a 'fingerprint' of the effects of chronic stress in several studies examining a variety of psychosocial stressors[39,107].

T cells and resilience to depression. Some of the most intriguing data regarding the role of the immune system in depression come from studies showing that T cells may protect against stress and depression in laboratory animals. For example, the adoptive transfer of T cells from animals exposed to chronic social defeat stress led to an antidepressant behavioural phenotype in stress-naive mice, which was associated with decreased pro-inflammatory cytokines in serum, a shift towards a neuroprotective M2 phenotype in microglia and increased neurogenesis in the hippocampus[108]. Similar results have been reported following acute stress in mice, in which effector T cell migration to the choroid plexus as a result of glucocorticoid induction of intercellular adhesion molecule 1 (ICAM1) expression in the choroid plexus was associated with reduced anxiety-like behaviour[109]. Mice with impaired release of glucocorticoids in response to stress were anxiety prone[109]. Immunization of anxiety-prone animals with a CNS-specific antigen restored T cell trafficking to the brain during stress and reversed anxiety-like behaviour in association with increased neurogenesis[109]. Immunization with a CNS-specific antigen also blocked stress-induced depression in mice[110]. The mechanism by which T cells influence resilience is believed to be related to their production of IL-4 within the meningeal space. Through as yet uncharacterized pathways, IL-4 then stimulates astrocytes to produce BDNF, and also promotes the conversion of meningeal monocytes and macrophages from a pro-inflammatory M1 phenotype to a less inflammatory M2 phenotype[111]. The movement of T cells throughout the brain, including the meningeal space, has become an area of special interest with the recent description of a brain lymphatic system that heretofore had gone unrecognized[112]. Data also indicate that T_{Reg} cells may have a role in reducing inflammation and supporting neuronal integrity during stress[113]. Similar reports have characterized T cells that are activated by vagal nerve stimulation to produce acetylcholine, which can inhibit NF-κB activation by binding to the α7 subunit of the nicotinic acetylcholine receptor[114].

Relevant to depression, however, peripheral T cell trafficking in response to glucocorticoids has been shown to be impaired in patients with depression, possibly owing to glucocorticoid resistance as a result of genetically mediated (for example, *FKBP5*) or inflammasome-mediated mechanisms targeting the glucocorticoid receptor[46,115]. In addition, inflammatory cytokines and their signalling pathways, including p38 MAPK, have direct inhibitory effects on glucocorticoid receptor function[116]. Moreover, patients with depression have been shown to have increased numbers of peripheral blood myeloid-derived suppressor cells, which inhibit T cell function[117]. Of note,

Myeloid-derived suppressor cells
A heterogeneous population of cells of myeloid origin that rapidly expands during inflammation and can potently suppress T cell responses. They are now being explored as potential therapeutic targets to inhibit immune responses in autoimmune disease or transplant rejection.

© 2015 Macmillan Publishers Limited. All rights reserved

Box 2 | **Guidelines for anti-inflammatory clinical trials in depression**

Based on the animal and human literature on the effects of cytokines on the brain, the following guidelines can inform clinical trials designed to test the cytokine hypothesis of depression.

Inflammation only occurs in subgroups of patients with depression[30]. Clinical trials should enrich for patient populations with evidence of increased inflammation, particularly those identified by a C-reactive protein (CRP) >3 mg L^{-1}, which has been shown to characterize patients with depression with altered reward circuitry and increased basal ganglia glutamate, as well as those who have shown a response to anti-cytokine therapy[30,84,95].

Anti-inflammatory drugs may harm patients without increased inflammation. Inflammatory cytokines and the innate immune response have pivotal roles in synaptic plasticity, neurogenesis, long-term potentiation (which is a fundamental process in learning and memory) and possibly antidepressant response[123,128].

Primary behavioural outcome variables should include measures of anhedonia and anxiety. Neuroimaging studies coupled with studies administering a variety of inflammatory stimuli, including the inflammatory cytokine interferon-α, endotoxin and typhoid vaccination, have revealed that inflammation targets neurocircuits in the brain that regulate motivation and reward as well as anxiety, arousal and alarm[35]. In addition, these symptoms have been shown to respond to anti-cytokine therapy in limited studies.

Drugs that specifically target inflammatory cytokines and/or their signalling pathways are preferable. The majority of clinical trials to date have used anti-inflammatory drugs (non-steroidal anti-inflammatory agents and minocycline, a tetracycline antibiotic) that have several off-target effects making the extant data relevant to testing the cytokine hypothesis of depression difficult to interpret[31].

Target engagement must be established in the periphery and ultimately the brain. Protein and gene expression markers of inflammation in the peripheral blood can serve as relevant proxies for inflammation in the brain[129], especially given evidence of trafficking of activated peripheral immune cells to the brain in stress-induced animal models of depression. Relevant therapeutic interventions should decrease peripheral inflammatory markers in concert with improvement of specific depressive symptoms. Translocator protein neuroimaging ligands may ultimately serve as direct measures of neuroinflammation and its inhibition by anti-inflammatory therapies in future clinical trials[61].

activation of the NLRP3 inflammasome leads to increased accumulation of myeloid-derived suppressor cells[118]. Decreased numbers of peripheral blood T_{Reg} cells and reduced concentrations of anti-inflammatory cytokines in the blood, including TGFβ and IL-10, have also been reported in depression[119]. Thus, it appears that patients with depression may have impairments in neuroprotective and anti-inflammatory T cell responses.

These findings suggest that therapies that boost such T cell responses could be used in patients with depression. Examples include immunization strategies (with CNS antigens as discussed above) that attract T cells to the brain or administration of bacteria, such as *Mycobacterium vaccae,* or parasites that stimulate T_{Reg} cell responses or T cell production of IL-4 (REFS 14,109,110,120). Indeed, colonization of pregnant dams with helminths attenuated the increase of hippocampal IL-1β in neonatal rats infected with bacteria and protected these animals from the subsequent development of microglial sensitization and cognitive dysfunction in adulthood. This effect was associated with increased *ex vivo* production of IL-4 and decreased production of IL-1β and TNF by splenic macrophages in response to LPS stimulation[120]. Finally, vagus nerve stimulation could be used to induce anti-inflammatory acetylcholine-producing T cells[121]. Although many strategies exist to activate anti-inflammatory T cell responses including the induction of T_{Reg} cells by administration of mesenchymal stem cells[122], the majority of the approaches discussed above have proof-of-concept data in animal models of depression. Nevertheless, the clinical relevance of these approaches has yet to be determined by randomized clinical trials in patients with depression.

Cytokine hypothesis of depression
A theoretical framework that suggests that cytokines have a primary role in alterations of neurotransmitter metabolism, neuroendocrine function, neuroplasticity, neurocircuitry and behaviour in a subgroup of patients with depression and increased inflammation.

Translational considerations

Our increasing understanding of how inflammatory processes contribute to depression, combined with the growing frustration over the lack of discovery of new antidepressants, have stimulated interest in the possibility that various classes of anti-inflammatory medications or other anti-inflammatory strategies (as discussed above) may hold promise as novel 'all-purpose' antidepressants. Unfortunately, it appears that anti-inflammatory agents may only demonstrate effective antidepressant activity in subgroups of patients who show evidence of increased peripheral inflammation, for example individuals with medical conditions including osteoarthritis and psoriasis that are characterized by increased levels of peripheral inflammation and patients with depression with increased inflammatory markers[29,30]. Moreover, in patients with depression who do not show elevated peripheral levels of inflammation, anti-inflammatory treatments may actually impair placebo responses that contribute to the effectiveness of all known antidepressant modalities[123]. In the only study to date examining the antidepressant effect of a cytokine antagonist in medically healthy adults with treatment-resistant depression, post hoc analysis revealed a dose-response relationship between baseline levels of peripheral inflammation and subsequent antidepressant response to the TNF inhibitor infliximab[30]. In patients with baseline plasma CRP concentrations ≥5 mg L^{-1}, infliximab outperformed placebo with an effect size similar to that observed in studies of standard antidepressants. Patients with a CRP >3 mg L^{-1}, the standard cut-off for high inflammation, also exhibited separation from placebo. Of note, this latter finding along with data demonstrating the relevance of a CRP >3 mg L^{-1} to altered reward circuitry and glutamate metabolism in depression as well as the prediction of subsequent depressive episodes (described above) aligns well with other diseases in which a CRP >3 mg L^{-1} is relevant to prediction and pathology including cardiovascular disease and diabetes. These data suggest that the cut-off for high inflammation in depression may be consistent with other disorders (BOX 2). Importantly, however, in patients with lower levels of inflammation, blockade of TNF with infliximab actually impaired the placebo response[30], suggesting that anti-inflammatory treatments in patients without inflammation may be detrimental, highlighting the growing recognition that the immune system has an important role in several processes central to neuronal integrity.

We conclude by offering the balanced perspective that anti-inflammatory therapies are unlikely to be all-purpose antidepressants. Perhaps we only think of standard antidepressants as all-purpose agents because we have never succeeded in developing predictive biomarkers that would reliably inform us of who will respond to any given agent.

© 2015 Macmillan Publishers Limited. All rights reserved

If so, then we view these agents as all-purpose, not because it is true but out of hope and ignorance. Thus, instead of being a negative, perhaps the finding that baseline inflammatory biomarkers such as CRP can predict subsequent symptomatic response to anti-inflammatory strategies is, in fact, the most positive development thus far in our quest to understand how the immune system might be harnessed to improve the treatment of depression.

Conclusion

In ancestral times, integration of inflammatory responses and behaviours of avoidance and alarm provided an evolutionary advantage in managing the microbial world. In the absence of the temporizing influence of commensal organisms that were rife in environments in which humans evolved, the inflammatory bias of the human species in the civilized world has been increasingly engaged in the complex world of psychosocial interactions and the inevitable stress it engenders. Responding to these sterile insults with activation of the inflammasome and mobilization of myeloid cells to the brain, the resultant release of inflammatory cytokines impinges on neurotransmitters and neurocircuits to lead to behaviours that are poorly suited for functioning in modern society. This inevitability of our evolutionary past is apparent in the high rates of depression that are seen in society today. There is also an increasing recognition of mechanisms of resilience that derive from our emerging understanding of the neuroprotective effects of a variety of T cell responses ranging from effector T cells that produce IL-4 to T_{Reg} cells with anti-inflammatory properties. A better understanding of these neuroprotective pathways and of the inflammatory mechanisms — from inflammasome activation to cell trafficking to the brain — that operate in patients with depression may lead to the development of novel anti-depressant therapies.

1. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **386**, 743–800 (2015).
2. Rush, A. J. *et al.* Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am. J. Psychiatry* **163**, 1905–1917 (2006).
3. Pace, T. W. *et al.* Increased stress-induced inflammatory responses in male patients with major depression and increased early life stress. *Am. J. Psychiatry* **163**, 1630–1633 (2006).
4. Bierhaus, A. *et al.* A mechanism converting psychosocial stress into mononuclear cell activation. *Proc. Natl Acad. Sci. USA* **100**, 1920–1925 (2003). **This study is one of the first demonstrations that a psychological stressor could activate fundamental inflammatory signalling pathways (that is, NF-κB) in human peripheral blood mononuclear cells.**
5. Aschbacher, K. *et al.* Maintenance of a positive outlook during acute stress protects against pro-inflammatory reactivity and future depressive symptoms. *Brain Behav. Immun.* **26**, 346–352 (2012).
6. Raison, C. L. & Miller, A. H. The evolutionary significance of depression in Pathogen Host Defense (PATHOS-D). *Mol. Psychiatry* **18**, 15–37 (2013). **This theoretical treatise proposes that depression, rather than being a maladaptive response to psychosocial challenge, is the outgrowth of an evolutionary advantage provided by a crosstalk between the immune system and the brain to survive ancestral challenges from pathogens and predators.**
7. Watson, P. J. & Andrews, P. W. Toward a revised evolutionary adaptationist analysis of depression: the social navigation hypothesis. *J. Affect. Disord.* **72**, 1–14 (2002).
8. Kinney, D. K. & Tanaka, M. An evolutionary hypothesis of depression and its symptoms, adaptive value, and risk factors. *J. Nerv. Ment. Dis.* **197**, 561–567 (2009).
9. Slavich, G. M. & Irwin, M. R. From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. *Psychol. Bull.* **140**, 774–815 (2014).
10. Seedat, S. *et al.* Cross-national associations between gender and mental disorders in the World Health Organization World Mental Health Surveys. *Arch. Gen. Psychiatry* **66**, 785–795 (2009).
11. Moieni, M. *et al.* Sex differences in depressive and socioemotional responses to an inflammatory challenge: implications for sex differences in depression. *Neuropsychopharmacology* **40**, 1709–1716 (2015).
12. Udina, M. *et al.* Interferon-induced depression in chronic hepatitis C: a systematic review and meta-analysis. *J. Clin. Psychiatry* **73**, 1128–1138 (2012).
13. Raison, C. L., Lowry, C. A. & Rook, G. A. Inflammation, sanitation, and consternation: loss of contact with coevolved, tolerogenic microorganisms and the pathophysiology and treatment of major depression. *Arch. Gen. Psychiatry* **67**, 1211–1224 (2010).
14. Rook, G. A., Lowry, C. A. & Raison, C. L. Hygiene and other early childhood influences on the subsequent function of the immune system. *Brain Res.* **1617**, 47–62 (2015).
15. Yirmiya, R. *et al.* Illness, cytokines, and depression. *Ann. NY Acad. Sci.* **917**, 478–487 (2000).
16. Miller, A. H., Maletic, V. & Raison, C. L. Inflammation and its discontents: the role of cytokines in the pathophysiology of major depression. *Biol. Psychiatry* **65**, 732–741 (2009).
17. Maes, M. Major depression and activation of the inflammatory response system. *Adv. Exp. Med. Biol.* **461**, 25–46 (1999).
18. Brambilla, P. *et al.* Increased M1/decreased M2 signature and signs of Th1/Th2 shift in chronic patients with bipolar disorder, but not in those with schizophrenia. *Transl Psychiatry* **4**, e406 (2014).
19. Drago, A., Crisafulli, C., Calabro, M. & Serretti, A. Enrichment pathway analysis. The inflammatory genetic background in bipolar disorder. *J. Affect Disord.* **179**, 88–94 (2015).
20. Mostafavi, S. *et al.* Type I interferon signaling genes in recurrent major depression: increased expression detected by whole-blood RNA sequencing. *Mol. Psychiatry* **19**,1267–1274 (2013).
21. Maes, M. Evidence for an immune response in major depression: a review and hypothesis. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **19**, 11–38 (1995). **In contrast to previous theories that primarily focused on reduced T cell responses to mitogenic stimuli, this is one of the first papers to posit that an activated immune system may have a role in the aetiology of depression.**
22. Bufalino, C., Hepgul, N., Aguglia, E. & Pariante, C. M. The role of immune genes in the association between depression and inflammation: a review of recent clinical studies. *Brain Behav. Immun.* **31**, 31–47 (2012).
23. Capuron, L. *et al.* Neurobehavioral effects of interferon-α in cancer patients: phenomenology and paroxetine responsiveness of symptom dimensions. *Neuropsychopharmacology* **26**, 643–652 (2002).
24. Reichenberg, A. *et al.* Cytokine-associated emotional and cognitive disturbances in humans. *Arch. Gen. Psychiatry* **58**, 445–452 (2001).
25. Bonaccorso, S. *et al.* Increased depressive ratings in patients with hepatitis C receiving interferon-α-based immunotherapy are related to interferon-α-induced changes in the serotonergic system. *J. Clin. Psychopharmacol.* **22**, 86–90 (2002).
26. Harrison, N. A. *et al.* Inflammation causes mood changes through alterations in subgenual cingulate activity and mesolimbic connectivity. *Biol. Psychiatry* **66**, 407–414 (2009).
27. Tyring, S. *et al.* Etanercept and clinical outcomes, fatigue, and depression in psoriasis: double-blind placebo-controlled randomised phase III trial. *Lancet* **367**, 29–35 (2006).
28. Abbott, R. *et al.* Tumour necrosis factor-α inhibitor therapy in chronic physical illness: a systematic review and meta-analysis of the effect on depression and anxiety. *J. Psychosom. Res.* **79**, 175–84 (2015).
29. Kohler, O. *et al.* Effect of anti-inflammatory treatment on depression, depressive symptoms, and adverse effects: a systematic review and meta-analysis of randomized clinical trials. *JAMA Psychiatry* **71**, 1381–1391 (2014).
30. Raison, C. L. *et al.* A randomized controlled trial of the tumor necrosis factor antagonist infliximab for treatment-resistant depression: the role of baseline inflammatory biomarkers. *JAMA Psychiatry* **70**, 31–41 (2013). **This report describes the results of the first double-blind, placebo-controlled trial of a monoclonal antibody against TNF to treat major depression, indicating that only patients with depression that have high levels of inflammation respond to cytokine antagonism.**
31. Miller, A. H. & Raison, C. L. Are anti-inflammatory therapies viable treatments for psychiatric disorders?: where the rubber meets the road. *JAMA Psychiatry* **72**, 527–528 (2015).
32. Michopoulos, V. *et al.* Association of *CRP* genetic variation and CRP level with elevated PTSD symptoms and physiological responses in a civilian population with high levels of trauma. *Am. J. Psychiatry* **172**, 353–362 (2015).
33. Fernandes, B. S. *et al.* C-reactive protein is increased in schizophrenia but is not altered by antipsychotics: meta-analysis and implications. *Mol. Psychiatry* http://dx.doi.org/10.1038/mp.2015.87 (2015).
34. Morris, S. E. & Cuthbert, B. N. Research Domain Criteria: cognitive systems, neural circuits, and dimensions of behavior. *Dialogues Clin. Neurosci.* **14**, 29–37 (2012).
35. Miller, A. H., Haroon, E., Raison, C. L. & Felger, J. C. Cytokine targets in the brain: impact on neurotransmitters and neurocircuits. *Depress. Anxiety* **30**, 297–306 (2013).
36. Cattaneo, A. *et al.* Candidate genes expression profile associated with antidepressants response in the GENDEP study: differentiating between baseline 'predictors' and longitudinal 'targets'. *Neuropsychopharmacology* **38**, 377–385 (2013).
37. Eurelings, L. S., Richard, E., Eikelenboom, P., van Gool, W. A. & Moll van Charante, E. P. Low-grade inflammation differentiates between symptoms of apathy and depression in community-dwelling older individuals. *Int. Psychogeriatr.* **27**, 639–647 (2015).
38. Pearson, T. A. *et al.* Markers of inflammation and cardiovascular disease: application to clinical and public health practice: a statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation* **107**, 499–511 (2003).

© 2015 Macmillan Publishers Limited. All rights reserved

39. Irwin, M. R. & Cole, S. W. Reciprocal regulation of the neural and innate immune systems. *Nat. Rev. Immunol.* **11**, 625–632 (2011).
40. Iwata, M., Ota, K. T. & Duman, R. S. The inflammasome: pathways linking psychological stress, depression, and systemic illnesses. *Brain Behav. Immun.* **31**, 105–114 (2013).
41. Strowig, T., Henao-Mejia, J., Elinav, E. & Flavell, R. Inflammasomes in health and disease. *Nature* **481**, 278–286 (2012).
42. Fleshner, M. Stress-evoked sterile inflammation, danger associated molecular patterns (DAMPs), microbial associated molecular patterns (MAMPs) and the inflammasome. *Brain Behav. Immun.* **27**, 1–7 (2013).
43. Cox, S. S. *et al.* Adrenergic and glucocorticoid modulation of the sterile inflammatory response. *Brain Behav. Immun.* **36**, 183–192 (2014).
44. Pan, Y., Chen, X. Y., Zhang, Q. Y. & Kong, L. D. Microglial NLRP3 inflammasome activation mediates IL-1β-related inflammation in prefrontal cortex of depressive rats. *Brain Behav. Immun.* **41**, 90–100 (2014).
45. Zhang, Y. *et al.* NLRP3 inflammasome mediates chronic mild stress-induced depression in mice via neuroinflammation. *Int. J. Neuropsychopharmacol.* **18**, pii: pyv006 (2015).
46. Paugh, S. W. *et al.* NALP3 inflammasome upregulation and CASP1 cleavage of the glucocorticoid receptor cause glucocorticoid resistance in leukemia cells. *Nat. Genet.* **47**, 607–614 (2015).
47. Rhen, T. & Cidlowski, J. A. Antiinflammatory action of glucocorticoids—new mechanisms for old drugs. *N. Engl. J. Med.* **353**, 1711–1723 (2005).
48. Raison, C. L. & Miller, A. H. When not enough is too much: the role of insufficient glucocorticoid signaling in the pathophysiology of stress-related disorders. *Am. J. Psychiatry* **160**, 1554–1565 (2003).
49. Pace, T. W., Hu, F. & Miller, A. H. Cytokine-effects on glucocorticoid receptor function: relevance to glucocorticoid resistance and the pathophysiology and treatment of major depression. *Brain Behav. Immun.* **21**, 9–19 (2007).
50. Alcocer-Gomez, E. *et al.* NLRP3 inflammasome is activated in mononuclear blood cells from patients with major depressive disorder. *Brain Behav. Immun.* **36**, 111–117 (2014).
This paper provides the first indication that activation of the inflammasome may contribute to elevated levels of inflammatory cytokines such as IL-1β and IL-18 in major depression, consistent with studies in laboratory animal models of depression that demonstrated that inhibition of NLRP3 can block the development of stress-induced depressive-like behaviour.
51. Stertz, L. *et al.* Damage-associated molecular patterns and immune activation in bipolar disorder. *Acta Psychiatr. Scand.* **132**, 211–217 (2015).
52. Rawdin, B. J. *et al.* Dysregulated relationship of inflammation and oxidative stress in major depression. *Brain Behav. Immun.* **31**, 143–152 (2013).
53. Maes, M., Galecki, P., Chang, Y. S. & Berk, M. A review on the oxidative and nitrosative stress (O&NS) pathways in major depression and their possible contribution to the (neuro)degenerative processes in that illness. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **35**, 676–692 (2011).
54. Mayer, E. A., Knight, R., Mazmanian, S. K., Cryan, J. F. & Tillisch, K. Gut microbes and the brain: paradigm shift in neuroscience. *J. Neurosci.* **34**, 15490–15496 (2014).
55. Maslanik, T. *et al.* Commensal bacteria and MAMPs are necessary for stress-induced increases in IL-1β and IL-18 but not IL-6, IL-10 or MCP-1. *PLoS ONE* **7**, e50636 (2012).
56. Lyte, M., Vulchanova, L. & Brown, D. R. Stress at the intestinal surface: catecholamines and mucosa-bacteria interactions. *Cell Tissue Res.* **343**, 23–32 (2011).
57. Rao, J. S., Harry, G. J., Rapoport, S. I. & Kim, H. W. Increased excitotoxicity and neuroinflammatory markers in postmortem frontal cortex from bipolar disorder patients. *Mol. Psychiatry* **15**, 384–392 (2010).
58. Steiner, J. *et al.* Immunological aspects in the neurobiology of suicide: elevated microglial density in schizophrenia and depression is associated with suicide. *J. Psychiatr. Res.* **42**, 151–157 (2008).
59. Torres-Platas, S. G., Cruceanu, C., Chen, G. G., Turecki, G. & Mechawar, N. Evidence for increased microglial priming and macrophage recruitment in the dorsal anterior cingulate white matter of depressed suicides. *Brain Behav. Immun.* **42**, 50–59 (2014).
This study provides some of the most compelling evidence that neuroinflammation occurs in major depression by demonstrating that monocytes traffic to the brain of patients with depressive symptoms and assume a perivascular localization in association with chemoattractant molecules such as CCL2, which has been shown to attract monocytes to the brain in animal models of stress.
60. Nagy, C. *et al.* Astrocytic abnormalities and global DNA methylation patterns in depression and suicide. *Mol. Psychiatry* **20**, 320–328 (2015).
61. Setiawan, E. *et al.* Role of translocator protein density, a marker of neuroinflammation, in the brain during major depressive episodes. *JAMA Psychiatry* **72**, 268–275 (2015).
62. Hannestad, J. *et al.* The neuroinflammation marker translocator protein is not elevated in individuals with mild-to-moderate depression: a [^{11}C]PBR28 PET study. *Brain Behav. Immun.* **33**, 131–138 (2013).
63. Sandiego, C. M. *et al.* Imaging robust microglial activation after lipopolysaccharide administration in humans with PET. *Proc. Natl Acad. Sci. USA* **112**, 12468–12473 (2015).
64. Quan, N. & Banks, W. A. Brain-immune communication pathways. *Brain Behav. Immun.* **21**, 727–735 (2007).
65. D'Mello, C., Le, T. & Swain, M. G. Cerebral microglia recruit monocytes into the brain in response to tumor necrosis factor-α signaling during peripheral organ inflammation. *J. Neurosci.* **29**, 2089–2102 (2009).
66. Hennessy, E., Griffin, E. W. & Cunningham, C. Astrocytes are primed by chronic neurodegeneration to produce exaggerated chemokine and cell infiltration responses to acute stimulation with the cytokines IL-1β and TNF-α. *J. Neurosci.* **35**, 8411–8422 (2015).
67. Wohleb, E. S. *et al.* β-adrenergic receptor antagonism prevents anxiety-like behavior and microglial reactivity induced by repeated social defeat. *J. Neurosci.* **31**, 6277–6288 (2011).
68. Wohleb, E. S., Powell, N. D., Godbout, J. P. & Sheridan, J. F. Stress-induced recruitment of bone marrow-derived monocytes to the brain promotes anxiety-like behavior. *J. Neurosci.* **33**, 13820–13833 (2013).
References 67 and 68 present a series of experiments that demonstrate a cellular pathway by which cytokines signals can be transmitted to the brain via trafficking of monocytes from the bone marrow to the brain parenchyma, a process mediated by catecholamines and CCL2.
69. Krugel, U., Fischer, J., Radicke, S., Sack, U. & Himmerich, H. Antidepressant effects of TNF-α blockade in an animal model of depression. *J. Psychiatr. Res.* **47**, 611–616 (2013).
70. Arends, S. *et al.* Baseline predictors of response and discontinuation of tumor necrosis factor-α blocking therapy in ankylosing spondylitis: a prospective longitudinal observational cohort study. *Arthritis Res. Ther.* **13**, R94 (2011).
71. Gillespie, C. F., Garlow, S. J., Binder, E. B., Schatzberg, A. F. & Nemeroff, C. B. in *Textbook of Psychopharmacology* (eds Schatzberg, A. F. & Nemeroff, C. B.) 903–944 (America Psychiatric Publishing, 2009).
72. Zhu, C. B. *et al.* Interleukin-1 receptor activation by systemic lipopolysaccharide induces behavioral despair linked to MAPK regulation of CNS serotonin transporters. *Neuropsychopharmacology* **35**, 2510–2520 (2010).
73. Neurauter, G. *et al.* Chronic immune stimulation correlates with reduced phenylalanine turnover. *Curr. Drug Metab.* **9**, 622–627 (2008).
74. Felger, J. C. *et al.* Tyrosine metabolism during interferon-α administration: association with fatigue and CSF dopamine concentrations. *Brain Behav. Immun.* **31**, 153–160 (2013).
75. Maes, M., Leonard, B. E., Myint, A. M., Kubera, M. & Verkerk, R. The new '5-HT' hypothesis of depression: cell-mediated immune activation induces indoleamine 2,3-dioxygenase, which leads to lower plasma tryptophan and an increased synthesis of detrimental tryptophan catabolites (TRYCATs), both of which contribute to the onset of depression. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **35**, 702–721 (2011).
76. Raison, C. L. *et al.* CSF concentrations of brain tryptophan and kynurenines during immune stimulation with IFN-α: relationship to CNS immune responses and depression. *Mol. Psychiatry* **15**, 393–403 (2010).
77. Steiner, J. *et al.* Severe depression is associated with increased microglial quinolinic acid in subregions of the anterior cingulate gyrus: evidence for an immune-modulated glutamatergic neurotransmission? *J. Neuroinflamm.* **8**, 94 (2011).
78. Tavares, R. G. *et al.* Quinolinic acid stimulates synaptosomal glutamate release and inhibits glutamate uptake into astrocytes. *Neurochem. Int.* **40**, 621–627 (2002).
79. Tilleux, S. & Hermans, E. Neuroinflammation and regulation of glial glutamate uptake in neurological disorders. *J. Neurosci. Res.* **85**, 2059–2070 (2007).
80. Hardingham, G. E., Fukunaga, Y. & Bading, H. Extrasynaptic NMDARs oppose synaptic NMDARs by triggering CREB shut-off and cell death pathways. *Nat. Neurosci.* **5**, 405–414 (2002).
81. Koo, J. W., Russo, S. J., Ferguson, D., Nestler, N. J. & Duman, R. S. Nuclear factor-κB is a critical mediator of stress-impaired neurogenesis and depressive behavior. *Proc. Natl Acad. Sci. USA* **107**, 2669–2674 (2010).
82. Goshen, I. *et al.* Brain interleukin-1 mediates chronic stress-induced depression in mice via adrenocortical activation and hippocampal neurogenesis suppression. *Mol. Psychiatry* **13**, 717–728 (2008).
83. Haroon, E. *et al.* IFN-α-induced cortical and subcortical glutamate changes assessed by magnetic resonance spectroscopy. *Neuropsychopharmacology* **39**, 1777–1785 (2014).
84. Haroon, E. *et al.* Conceptual convergence:increased inflammation is associated with increased basal ganglia glutamate in patients with major depression. *Mol. Psychiatry* http://dx.doi.org/10.1038/mp.2015.206 (2015).
85. Walker, A. K. *et al.* NMDA receptor blockade by ketamine abrogates lipopolysaccharide-induced depressive-like behavior in C57BL/6J mice. *Neuropsychopharmacology* **38**, 1609–1616 (2013).
86. O'Connor, J. C. *et al.* Lipopolysaccharide-induced depressive-like behavior is mediated by indoleamine 2,3-dioxygenase activation in mice. *Mol. Psychiatry* **14**, 511–522 (2009).
87. Duman, R. S. & Monteggia, L. M. A neurotrophic model for stress-related mood disorders. *Biol. Psychiatry* **59**, 1116–1127 (2006).
88. Hodes, G. E. *et al.* Individual differences in the peripheral immune system promote resilience versus susceptibility to social stress. *Proc. Natl Acad. Sci. USA* **111**, 16136–16141 (2014).
This paper gives compelling evidence in laboratory animals that individual differences in the behavioural response to a social stressor is mediated by individual differences in the production of the inflammatory cytokine IL-6, suggesting that genetic and/or environmental factors that regulate inflammatory responses can determine stress-induced depressive-like behaviour.
89. Capuron, L. *et al.* Dopaminergic mechanisms of reduced basal ganglia responses to hedonic reward during interferon-α administration. *Arch. Gen. Psychiatry* **69**, 1044–1053 (2012).
Using multimodal neuroimaging, this report provides an integration of the impact of inflammatory cytokines on reward pathways and dopamine metabolism in the basal ganglia that leads to alterations in behaviours relevant to motivation and ultimately anhedonia — a core symptom of depression.
90. Eisenberger, N. I. *et al.* Inflammation-induced anhedonia: endotoxin reduces ventral striatum responses to reward. *Biol. Psychiatry* **68**, 748–754 (2010).
91. Felger, J. C. *et al.* Chronic interferon-α decreases dopamine 2 receptor binding and striatal dopamine release in association with anhedonia-like behavior in nonhuman primates. *Neuropsychopharmacology* **38**, 2179–2187 (2013).
92. Dowell, N. G. *et al.* Acute changes in striatal microstructure predict the development of interferon-α induced fatigue. *Biol. Psychiatry* http://dx.doi.org/10.1016/j.biopsych.2015.05.015 (2015).
93. Harrison, N. A., Cercignani, M., Voon, V. & Critchley, H. D. Effects of inflammation on hippocampus and substantia nigra responses to novelty in healthy human participants. *Neuropsychopharmacology* **40**, 831–838 (2015).
94. Harrison, N. A. *et al.* A neuro-computational account of how inflammation enhances sensitivity to punishments versus rewards. *Biol. Psychiatry* http://dx.doi.org/10.1016/j.biopsych.2015.07.018 (2015).

© 2015 Macmillan Publishers Limited. All rights reserved

95. Felger, J. C. *et al.* Inflammation is associated with decreased functional connectivity within corticostriatal reward circuitry in depression. *Mol. Psychiatry* http://dx.doi.org/10.1038/mp.2015.168 (2015).
This neuroimaging study is the first to demonstrate that the impact of immune stimuli (for example, IFNα, typhoid vaccination and endotoxin) on the brains of otherwise non-depressed individuals extended to patients with depression, whose increased inflammation was found to reduce functional connectivity in reward-related neurocircuitry leading to anhedonia, a core symptom of depression.
96. Harrison, N. A. *et al.* Neural origins of human sickness in interoceptive responses to inflammation. *Biol. Psychiatry* **66**, 415–422 (2009).
97. Slavich, G. M., Way, B. M., Eisenberger, N. I. & Taylor, S. E. Neural sensitivity to social rejection is associated with inflammatory responses to social stress. *Proc. Natl Acad. Sci. USA* **107**, 14817–14822 (2010).
This manuscript provides one of the first and most detailed descriptions of the relationship between stress-induced activation of inflammatory responses and sensitivity to a psychosocial stressor, identifying key brain regions such as the dACC as a target of cytokines, leading to anxiety, arousal and alarm.
98. Eisenberger, N. I. & Lieberman, M. D. Why rejection hurts: a common neural alarm system for physical and social pain. *Trends Cogn. Sci.* **8**, 294–300 (2004).
99. Muscatell, K. A. *et al.* Greater amygdala activity and dorsomedial prefrontal-amygdala coupling are associated with enhanced inflammatory responses to stress. *Brain Behav. Immun.* **43**, 46–53 (2015).
100. Gimeno, D. *et al.* Associations of C-reactive protein and interleukin-6 with cognitive symptoms of depression: 12-year follow-up of the Whitehall II study. *Psychol. Med.* **39**, 413–423 (2009).
101. Au, B., Smith, K. J., Gariepy, G. & Schmitz, N. The longitudinal associations between C-reactive protein and depressive symptoms: evidence from the English Longitudinal Study of Ageing (ELSA). *Int. J. Geriatr. Psychiatry* **30**, 976–984 (2015).
102. Duivis, H. E. *et al.* Depressive symptoms, health behaviors, and subsequent inflammation in patients with coronary heart disease: prospective findings from the heart and soul study. *Am. J. Psychiatry* **168**, 913–920 (2011).
103. Baumeister, D., Akhtar, R., Ciufolini, S., Pariante, C. M. & Mondelli, V. Childhood trauma and adulthood inflammation: a meta-analysis of peripheral C-reactive protein, interleukin-6 and tumour necrosis factor-α. *Mol. Psychiatry* http://dx.doi.org/10.1038/mp.2015.67 (2015).
104. Danese, A., Pariante, C. M., Caspi, A., Taylor, A. & Poulton, R. Childhood maltreatment predicts adult inflammation in a life-course study. *Proc. Natl Acad. Sci. USA* **104**, 1319–1324 (2007).
105. Tartter, M., Hammen, C., Bower, J. E., Brennan, P. A. & Cole, S. Effects of chronic interpersonal stress exposure on depressive symptoms are moderated by genetic variation at IL6 and IL1β in youth. *Brain Behav. Immun.* **46**, 104–111 (2015).
106. Klengel, T. *et al.* Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nat. Neurosci.* **16**, 33–41 (2013).
107. Miller, G. E. *et al.* A functional genomic fingerprint of chronic stress in humans: blunted glucocorticoid and increased NF-κB signaling. *Biol. Psychiatry* **64**, 266–272 (2008).
108. Brachman, R. A., Lehmann, M. L., Maric, D. & Herkenham, M. Lymphocytes from chronically stressed mice confer antidepressant-like effects to naive mice. *J. Neurosci.* **35**, 1530–1538 (2015).
109. Lewitus, G. M., Cohen, H. & Schwartz, M. Reducing post-traumatic anxiety by immunization. *Brain Behav. Immun.* **22**, 1108–1114 (2008).
110. Lewitus, G. M. *et al.* Vaccination as a novel approach for treating depressive behavior. *Biol. Psychiatry* **65**, 283–288 (2009).
References 109 and 110 provide the first evidence that engaging T cells to traffic to the brain during stress can block depressive- and anxiety-like behaviour and induce growth factors and neurogenesis in the brain, representing an entirely novel approach to the treatment of depression that takes advantage of the capacity of T cells to support neuronal integrity.
111. Derecki, N. C. *et al.* Regulation of learning and memory by meningeal immunity: a key role for IL-4. *J. Exp. Med.* **207**, 1067–1080 (2010).
112. Louveau, A. *et al.* Structural and functional features of central nervous system lymphatic vessels. *Nature* **523**, 337–341 (2015).
113. Kim, S. J. *et al.* $CD4^+CD25^+$ regulatory T cell depletion modulates anxiety and depression-like behaviors in mice. *PLoS ONE* **7**, e42054 (2012).
114. Rosas-Ballina, M. *et al.* Acetylcholine-synthesizing T cells relay neural signals in a vagus nerve circuit. *Science* **334**, 98–101 (2011).
115. Bauer, M. E. *et al.* Dexamethasone-induced effects on lymphocyte distribution and expression of adhesion molecules in treatment-resistant depression. *Psychiatry Res.* **113**, 1–15 (2002).
116. Wang, X., Wu, H. & Miller, A. H. Interleukin 1α (IL-1α) induced activation of p38 mitogen-activated protein kinase inhibits glucocorticoid receptor function. *Mol. Psychiatry* **9**, 65–75 (2004).
117. Wei, J., Zhang, M. & Zhou, J. Myeloid-derived suppressor cells in major depression patients suppress T-cell responses through the production of reactive oxygen species. *Psychiatry Res.* **228**, 695–701 (2015).
118. van Deventer, H. W. *et al.* The inflammasome component NLRP3 impairs antitumor vaccine by enhancing the accumulation of tumor-associated myeloid-derived suppressor cells. *Cancer Res.* **70**, 10161–10169 (2010).
119. Li, Y. *et al.* Altered expression of $CD4^+CD25^+$ regulatory T cells and its $5\text{-}HT_{1a}$ receptor in patients with major depression disorder. *J. Affect. Disord.* **124**, 68–75 (2010).
120. Williamson, L. L. *et al.* Got worms? Perinatal exposure to helminths prevents persistent immune sensitization and cognitive dysfunction induced by early-life infection. *Brain Behav. Immun.* (2015).
121. Olofsson, P. S., Rosas-Ballina, M., Levine, Y. A. & Tracey, K. J. Rethinking inflammation: neural circuits in the regulation of immunity. *Immunol. Rev.* **248**, 188–204 (2012).
122. Wang, Y., Chen, X., Cao, W. & Shi, Y. Plasticity of mesenchymal stem cells in immunomodulation: pathological and therapeutic implications. *Nat. Immunol.* **15**, 1009–1016 (2014).
123. Warner-Schmidt, J. L., Vanover, K. E., Chen, E. Y., Marshall, J. J. & Greengard, P. Antidepressant effects of selective serotonin reuptake inhibitors (SSRIs) are attenuated by antiinflammatory drugs in mice and humans. *Proc. Natl Acad. Sci. USA* **108**, 9262–9267 (2011).
124. Gurven, M. & Kaplan, H. Longevity among hunter-gatherers: a cross-cultural examination. *Popul. Dev. Rev.* **33**, 321–365 (2007).
125. Fumagalli, M. *et al.* Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet.* **7**, e1002355 (2011).
126. Kuningas, M. *et al.* Selection for genetic variation inducing pro-inflammatory responses under adverse environmental conditions in a Ghanaian population. *PLoS ONE* **4**, e7795 (2009).
127. Ratcliffe, M. A bad case of the flu?: the comparative phenomenology of depression and somatic illness. *J. Conscious. Studies* **20**, 198–218 (2013).
128. Marin, I. & Kipnis, J. Learning and memory… and the immune system. *Learn. Mem.* **20**, 601–606 (2013).
129. Mehta, D. *et al.* Transcriptional signatures related to glucose and lipid metabolism predict treatment response to the tumor necrosis factor antagonist infliximab in patients with treatment-resistant depression. *Brain Behav. Immun.* **31**, 205–215 (2013).

Competing interests statement
The authors declare no competing interests.

© 2015 Macmillan Publishers Limited. All rights reserved

Cell Metabolism, Volume *15*

Supplemental Information

Time-Restricted Feeding without Reducing Caloric Intake Prevents Metabolic Diseases in Mice Fed a High-Fat Diet

Megumi Hatori, Christopher Vollmers, Amir Zarrinpar, Luciano DiTacchio, Eric A. Bushong, Shubhroz Gill, Mathias Leblanc, Amandine Chaix, Matthew Joens, James A.J. Fitzpatrick, Mark H. Ellisman, and Satchidananda Panda

Inventory of Supplemental Information

Figure S1. Representative pictures of mice under different feeding regimens. Relates to Figure 1.

Figure S2. Restoration of normal metabolic activity and circadian clock in tRF. Relates to Figure 2.

Figure S3. Temporal patterns of gene expression in mice fed a normal diet. Relates to Figures 3, 4 and 6.

Figure S4. Hepatic nucleotide metabolites, free fatty acids and pro-inflammatory fatty acid. Relates to Figures 3 and 4.

Figure S5. Temporal pattern of hepatic gene expression in mice fed a NC or HF diet under *ad lib* and tRF paradigms. Relates to Figures 5 and 6.

Figure S6. Bile acid levels are increased in the liver of FT mice compared to those found in the FA mice. Related to Figure 6.

Figure S7. Relative levels and diurnal accumulation pattern of key metabolites in the mouse liver. Relates to Figures 2-4, 6 and 7.

Table S1. Liver metabolites and their relative abundance in mice fed NC or HF diet under *ad lib* and tRF regimens. Relates to Figures 2-4, 6 and 7.

Table S2. Major subcellular structures and their percent volume fraction in hepatocytes from FA and FT mice as determined by serial block face scanning electron microscopy of liver samples. Relates to Figure 5D.

Supplementary Figure 1

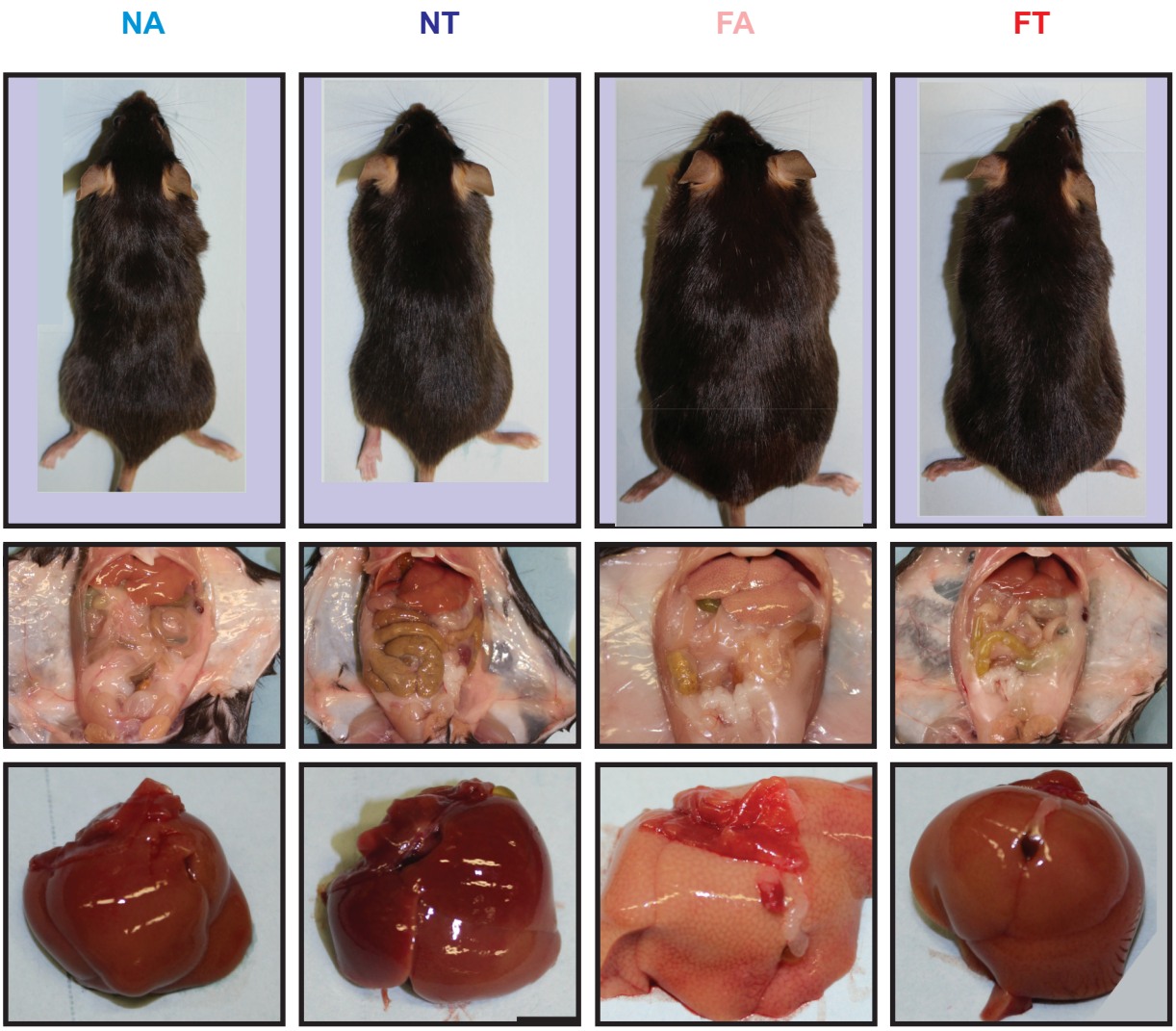


Figure S1. Pictures of randomly selected whole body (top), body cavity (middle) and liver (bottom) of mice on NA, NT, FA and FT regimens. Mice were sacrificed after 18 weeks on the feeding regimen. Relates to Figure 1.

Supplementary Figure 2

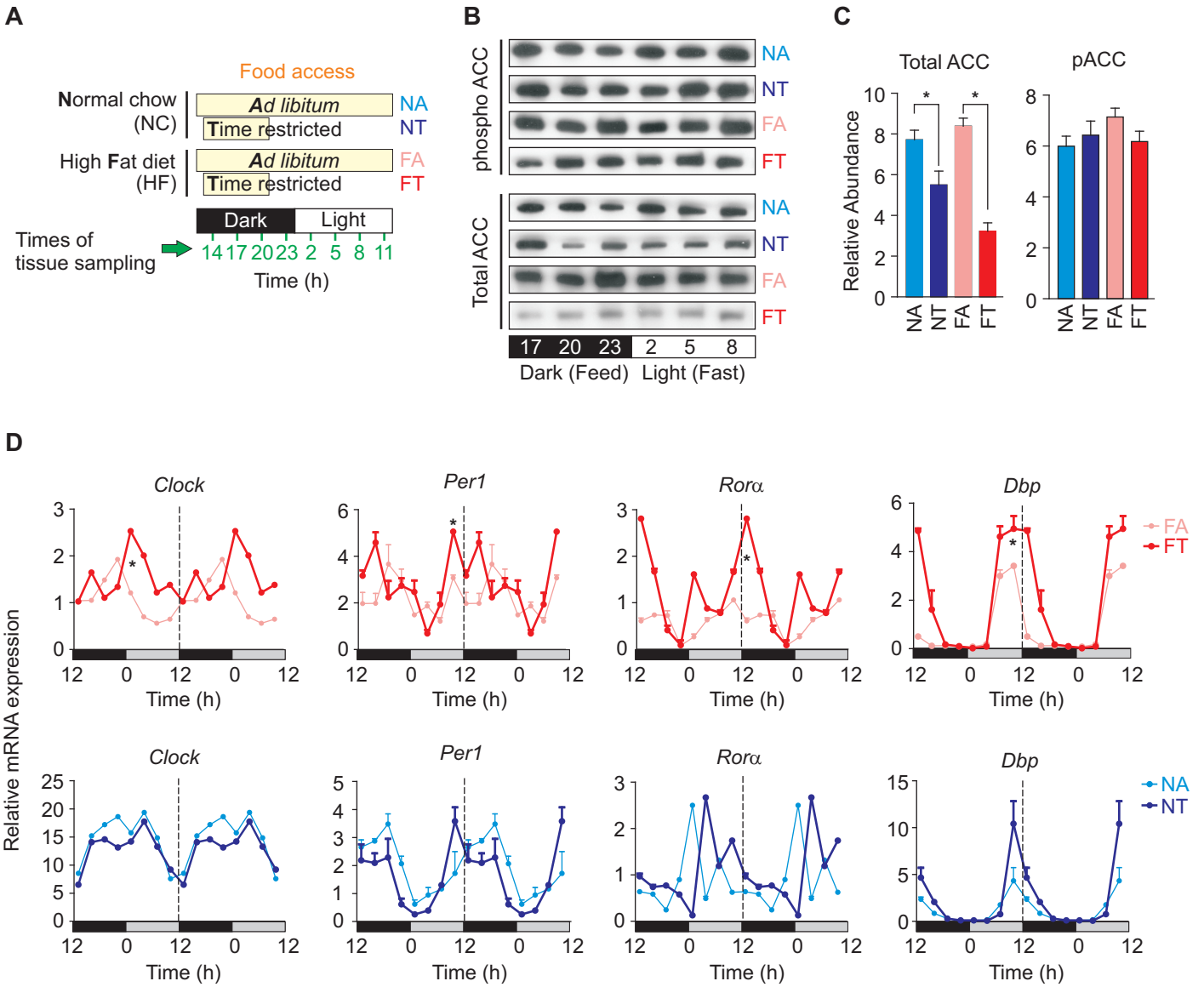


Figure S2. Time restricted feeding restored normal activities of metabolic regulators and circadian rhythms in mice fed a high fat diet. (A) Schematic representation of time of tissue samples collected for RNA, protein and metabolite analyses. At each time point, 4 mice under each feeding regimen were sacrificed, and their tissues collected and flash frozen for subsequent analyses. Each frozen tissue sample was pulverized to a fine powder and split for RNA extraction, protein extraction and metabolite analyses. For immuno blot analyses shown in Figures 2A, 2B and S2B, liver samples from three time points corresponding to the time of food access in tRF mice and three time points that are 12 h away during the day time were analyzed. **(B)** Immunoblot analyses of phospho-ACC (pACC) and total ACC from the liver of four groups of mice during nighttime feeding or daytime fasting. Equal quantities of protein were loaded in each lane. No apparent diurnal change in ACC or pACC was detected. **(C)** Densitometric quantification of average (+ SEM, n = 6) levels of total ACC and pACC in the livers of four groups of mice. The reduction in total ACC in the liver of mice under FT regimen is likely due to the increased expression of its repressor REV-ERBα (Feng et al., 2011). **(D)** Double-plotted mean (+ SEM, n = 4) mRNA levels of *Clock*, *Per1*, *Rorα* and *Dbp* from the liver show improved oscillation under tRF paradigm. The levels of expression and/or the peak:trough ratios are remarkably increased in FT relative to FA liver, while the changes are modest between NA and NT groups. Expression levels were measured by qRT-PCR and normalized against *Gapdh* mRNA. Results shown here relate to Figure 2.

Supplementary Figure 3

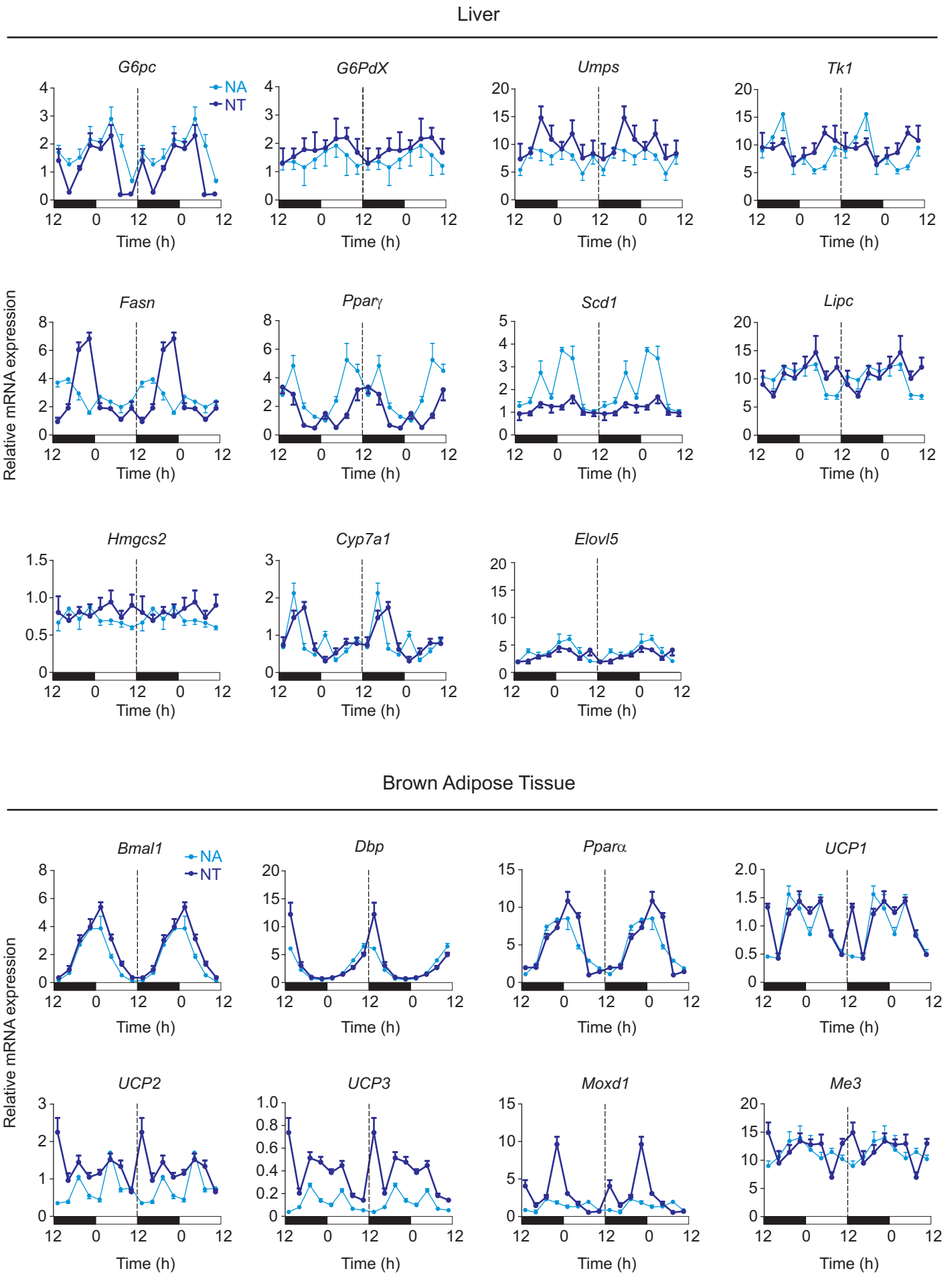


Figure S3. Temporal patterns of gene expression in mice fed a normal diet. Double-plotted average (+ SEM, n = 4) mRNA levels of genes regulating metabolism of glucose, nucleotides, fatty acids, cholesterol and bile acids in the liver or of circadian oscillator components, regulators of fat metabolism and mitochondria function in the BAT of NA and NT mice at different times of the day. Transcript levels were measured by qRT-PCR, and normalized to *Gapdh* RNA levels. Broken line separates double plotted data. Results relate to Figures 3, 4, and 6.

Supplementary Figure 4

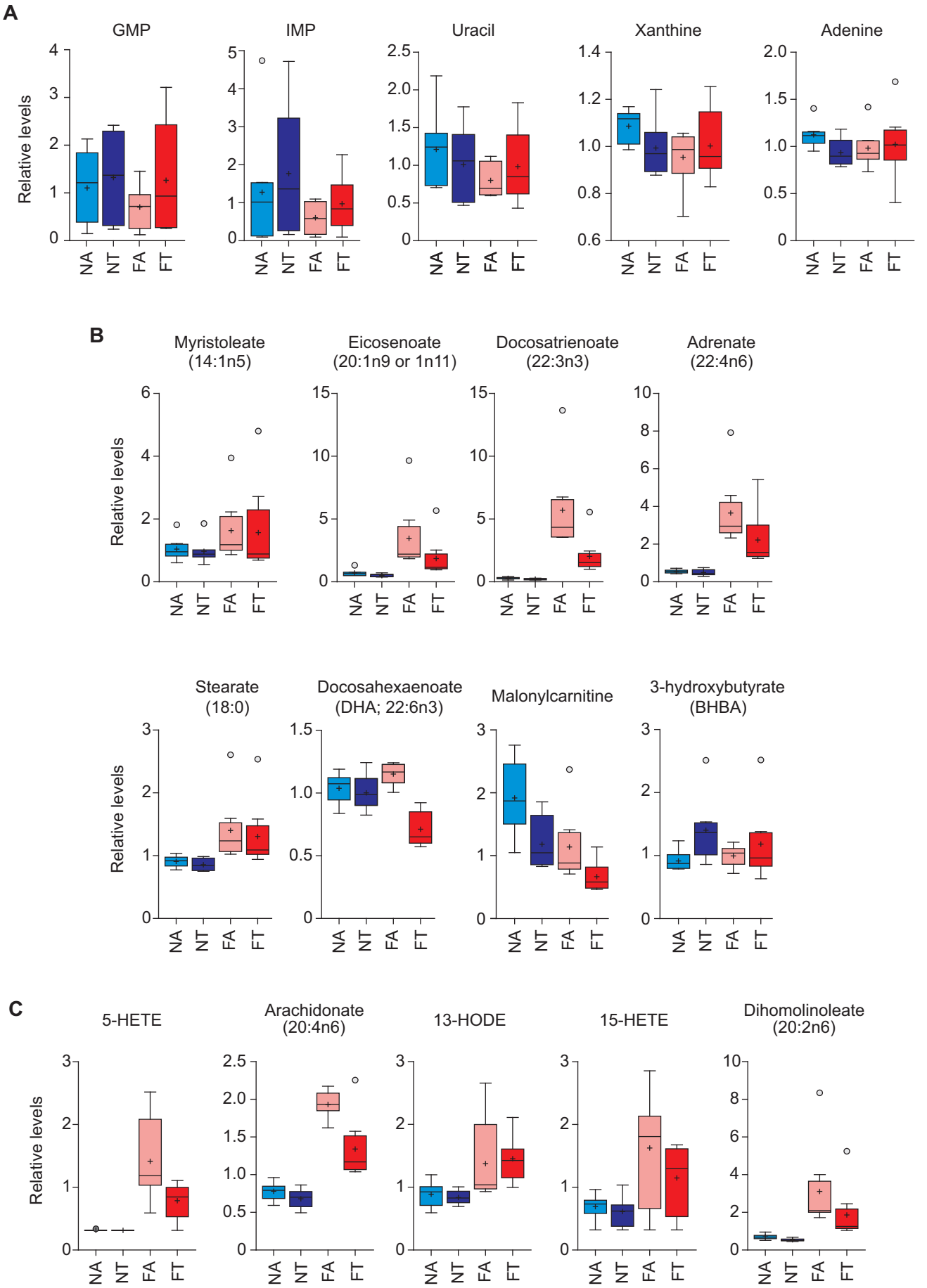
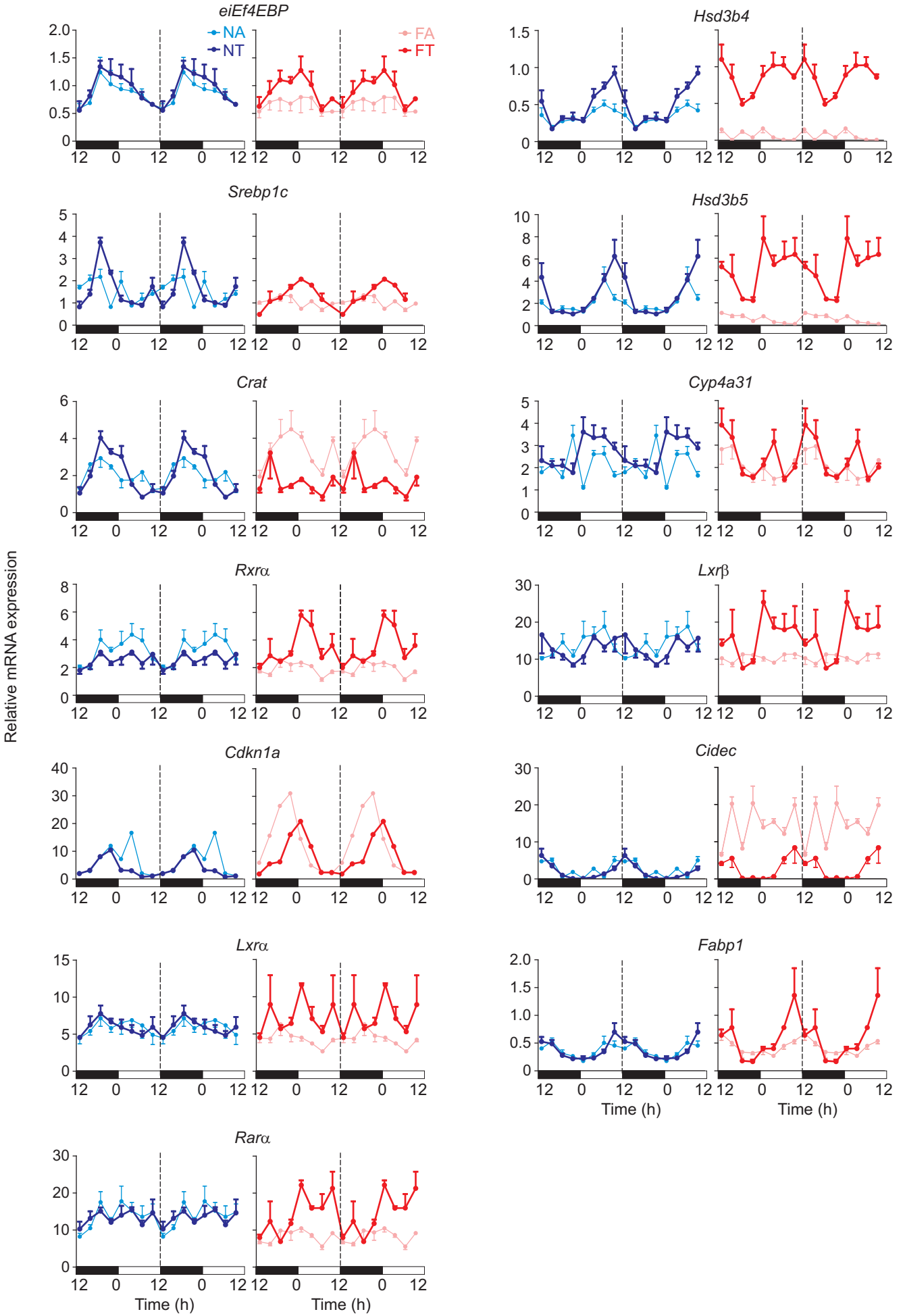


Figure S4. Hepatic accumulation of nucleotide metabolites, free fatty acids and pro-inflammatory fatty acid metabolites are modulated under tRF paradigm. (A) Levels of several intermediates of nucleotide metabolism are increased in the liver of FT mice relative to their levels in FA mice. This is consistent with the changes in gene expression and in nucleotide metabolites shown in Figures 3F-3H. **(B)** Relative levels of representative long chain free fatty acids are increased in mice fed high fat diet. The levels of the majority of long chain free fatty acids detected were reduced in the liver of FT mice. Mice fed normal chow generally contained small amount of free fatty acids. Nevertheless, the tRF paradigm further reduced the levels of hepatic free fatty acids. **(C)** Similar attenuation of pro-inflammatory fatty acids was observed in the liver. Results presented in Figure S4B and S4C relate to Figure 4. Results were plotted in whisker plots where the box denotes the middle 80 percentile with mean (+) and median (-). Error bars denote maximum and minimum of distribution with extreme data points marked (o).

Supplementary Figure 5



Supplementary Figure 5 (continued)

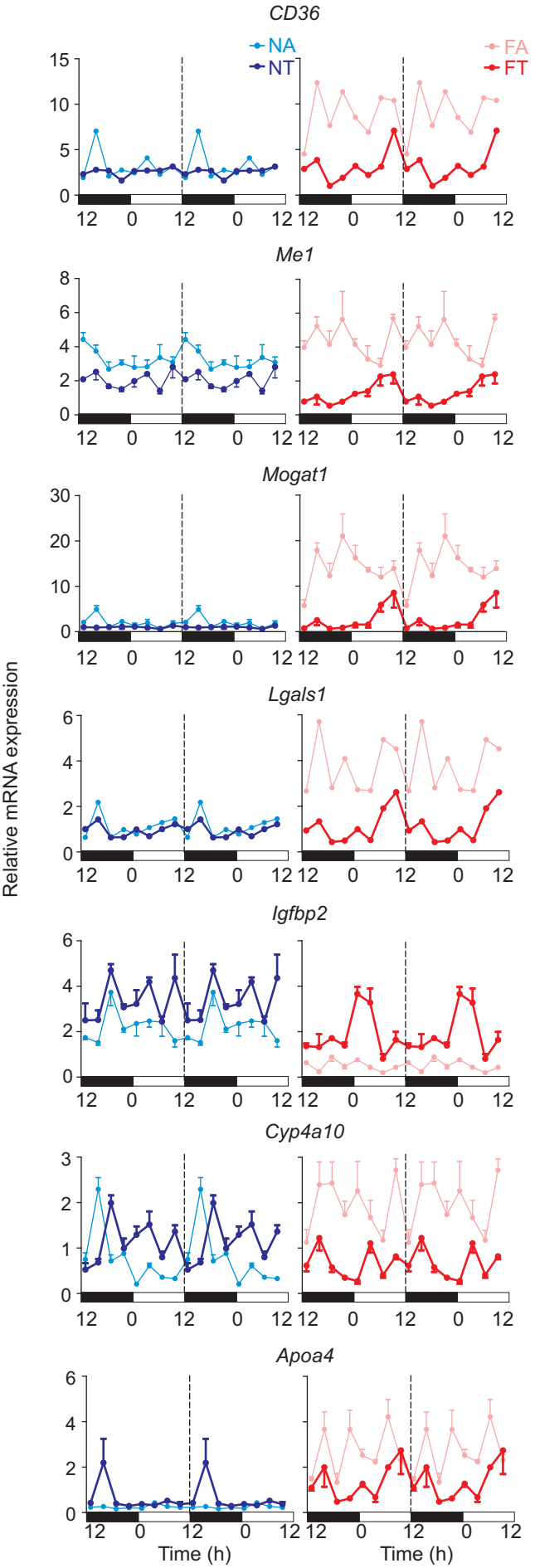


Figure S5. Temporal hepatic gene expression pattern in mice fed a NC or HF diet under *ad lib* and tRF paradigms. Double-plotted average (+ SEM, n = 4) mRNA levels of genes implicated in protein translation, steroid metabolism and in several liver diseases are shown. Transcript levels were measured by qRT-PCR, normalized to *Gapdh* RNA levels. Broken line separates double plotted data. Note that the overall changes in gene expression upon time restricted feeding are moderate in mice fed a normal diet, while temporal feeding pattern has a pronounced effect when mice are fed a high fat diet. Results relate to Figures 5 and 6.

Supplementary Figure 6

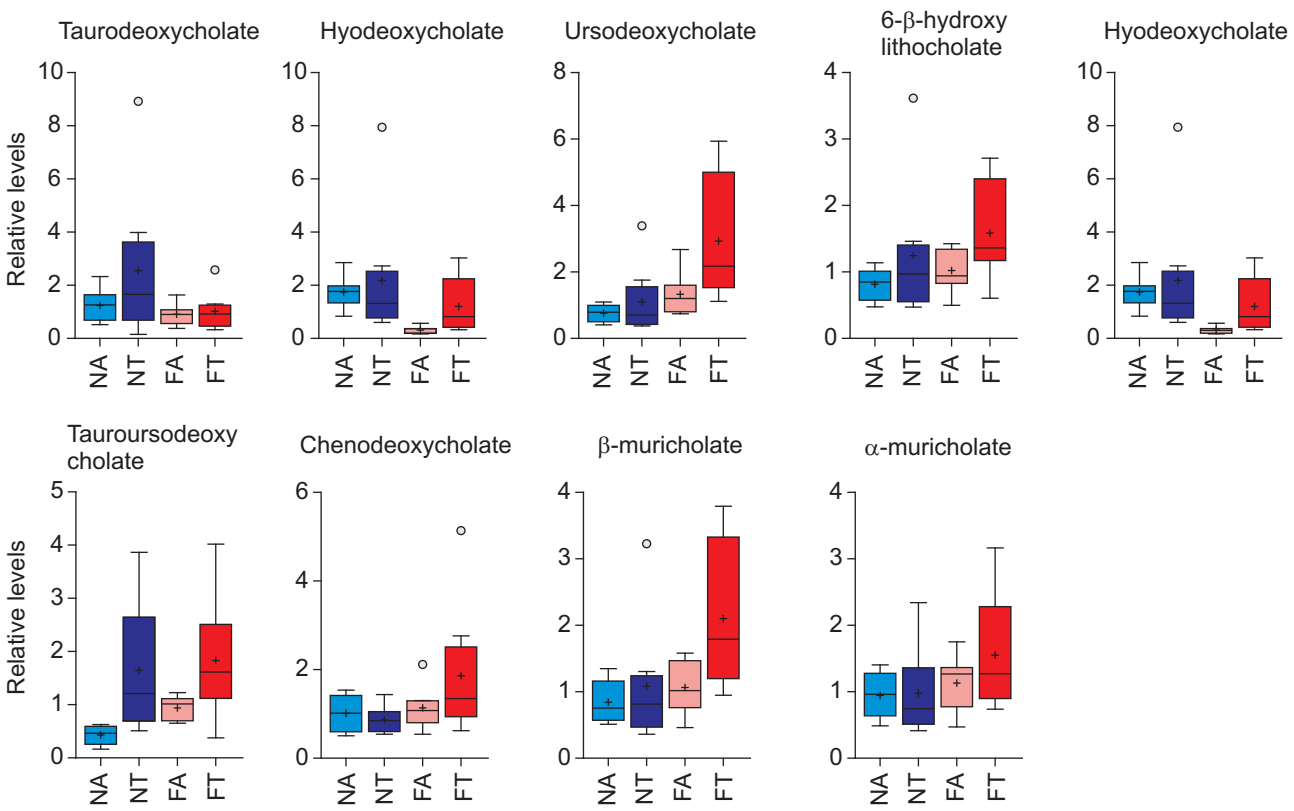


Figure S6. Bile acid levels are increased in the liver of FT mice compared to those found in the FA mice. Results were plotted in whisker plots where the box denotes the middle 80 percentile with mean (+) and median (-). Error bars denote maximum and minimum of distribution with extreme data points marked (o). Related to Figure 6.

Supplementary Figure 7

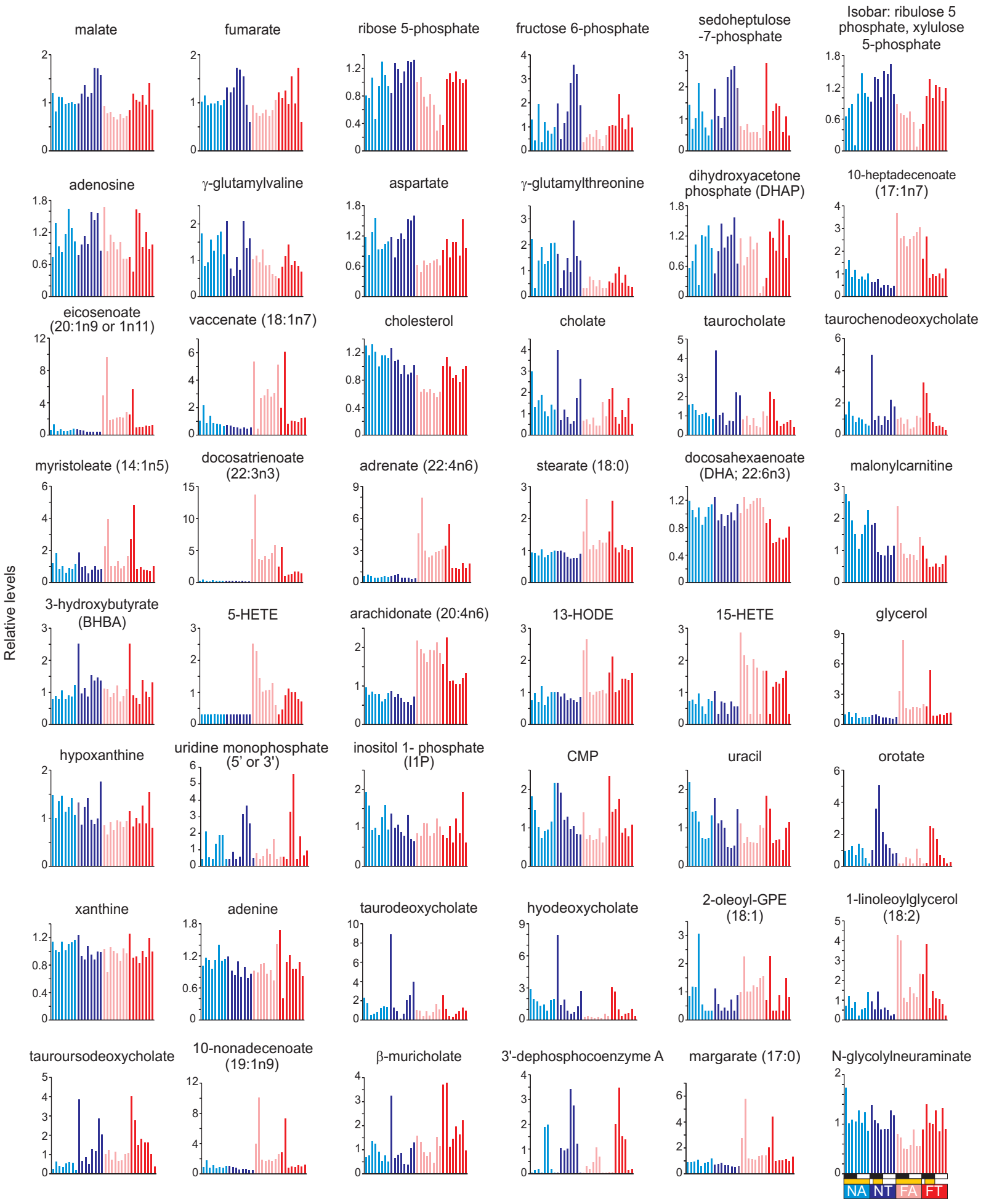


Figure S7. **Relative levels and the diurnal accumulation pattern of key liver metabolites in the mouse liver are defined by nutrient type and time of feeding.** Normalized levels of several metabolites at 8 different time points spaced 3 h apart over one full day are shown for each of the four feeding groups (NA, NT, FA and FT). Times of measurement relative to light:dark cycle and periods of food access are shown at the bottom right panel. Values shown here relate to the whisker plots of metabolites shown in Figures 2-4, 6 and 7. Relative values and associated statistical parameters are shown in Table S1.

Table S2

	Percentage of volume	
	FT	FA
Mitochondrion	19.92%	8.24%
Cytosol	33.56%	24.19%
Vacuole	2.80%	16.05%
Nucleus	6.58%	2.60%
Capillaries	13.83%	8.57%
Glycogen	14.22%	5.75%
Lipid Body	0.58%	30.15%
Endoplasmic Reticulum	8.51%	4.45%

Table S2. Major subcellular structures and their percent volume fraction in hepatocytes from FA and FT mice as determined by serial block face scanning electron microscopy of liver samples. Relates to Figure 5D.

Supplemental Experimental Procedures

Histology and Electron Microscopy. Sections (6 µm) of formaldehyde fixed liver, WAT and BAT were stained following standard H&E. Liver sections were also stained using the Sirius Red method. H&E and Steatohepatitis was scored using a semi-quantitative method derived from Brunt *et al*. (Brunt et al., 2004) measuring the degree of steatosis (0-3), ballooning degeneration (0-2), lobular (0-3) and portal (0-2) inflammation and fibrosis (0-4). Histology sections were digitally imaged using a fixed, high-resolution Leica DFC420 digital camera mounted on a Leica DMLS microscope. A serial block-face scanning electron microscopy (SBFSEM) volume was collected from both FT and FA liver following protocols as previously described (Williams et al., 2011), using a Gatan 3View system installed on a Zeiss Sigma SEM. Volumes were 12,000 pixel x 12,000 pixel x 500 slices with voxel dimensions of 5.63 nm XY x 70 nm Z. Image analysis was performed using IMOD (http://bio3d.colorado.edu/imod/). Stereology was performed using a custom plug-in for IMOD.

Transcript, protein and metabolite analyses.

Anti-pACC, anti-ACC, anti-pS6 and anti-S6 antibodies were purchased from Cell Signaling Technologies (Cat No. 3661S, 3661S 4858S and 2217S, respectively). qRT-PCR primer sequence information is available upon request. For rhythmic transcripts, average expression (± SEM) from experimental replicates from each time point over 24 h are double plotted as a line chart.

Supplemental References

Brunt, E.M., Neuschwander-Tetri, B.A., Oliver, D., Wehmeier, K.R., and Bacon, B.R. (2004). Nonalcoholic steatohepatitis: histologic features and clinical correlations with 30 blinded biopsy specimens. Human pathology 35, 1070-1082.

Feng, D., Liu, T., Sun, Z., Bugge, A., Mullican, S.E., Alenghat, T., Liu, X.S., and Lazar, M.A. (2011). A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. Science 331, 1315-1319.

Williams, M.E., Wilke, S.A., Daggett, A., Davis, E., Otto, S., Ravi, D., Ripley, B., Bushong, E.A., Ellisman, M.H., Klein, G., and Ghosh, A. (2011). Cadherin-9 regulates synapse-specific differentiation in the developing hippocampus. Neuron 71, 640-655.

ONLINE FIRST

Toward a Neuroimaging Treatment Selection Biomarker for Major Depressive Disorder

Callie L. McGrath, BA; Mary E. Kelley, PhD; Paul E. Holtzheimer III, MD; Boadie W. Dunlop, MD; W. Edward Craighead, PhD; Alexandre R. Franco, PhD; R. Cameron Craddock, PhD; Helen S. Mayberg, MD

Importance: Currently, fewer than 40% of patients treated for major depressive disorder achieve remission with initial treatment. Identification of a biological marker that might improve these odds could have significant health and economic impact.

Objective: To identify a candidate neuroimaging "treatment-specific biomarker" that predicts differential outcome to either medication or psychotherapy.

Design: Brain glucose metabolism was measured with positron emission tomography prior to treatment randomization to either escitalopram oxalate or cognitive behavior therapy for 12 weeks. Patients who did not remit on completion of their phase 1 treatment were offered enrollment in phase 2 comprising an additional 12 weeks of treatment with combination escitalopram and cognitive behavior therapy.

Setting: Mood and anxiety disorders research program at an academic medical center.

Participants: Men and women aged 18 to 60 years with currently untreated major depressive disorder.

Intervention: Randomized assignment to 12 weeks of treatment with either escitalopram oxalate (10-20 mg/d) or 16 sessions of manual-based cognitive behavior therapy.

Main Outcome and Measure: Remission, defined as a 17-item Hamilton Depression Rating Scale score of 7 or less at both weeks 10 and 12, as assessed by raters blinded to treatment.

Results: Positive and negative predictors of remission were identified with a 2-way analysis of variance treatment (escitalopram or cognitive behavior therapy) × outcome (remission or nonresponse) interaction. Of 65 protocol completers, 38 patients with clear outcomes and usable positron emission tomography scans were included in the primary analysis: 12 remitters to cognitive behavior therapy, 11 remitters to escitalopram, 9 nonresponders to cognitive behavior therapy, and 6 nonresponders to escitalopram. Six limbic and cortical regions were identified, with the right anterior insula showing the most robust discriminant properties across groups (effect size = 1.43). Insula hypometabolism (relative to whole-brain mean) was associated with remission to cognitive behavior therapy and poor response to escitalopram, while insula hypermetabolism was associated with remission to escitalopram and poor response to cognitive behavior therapy.

Conclusions and Relevance: If verified with prospective testing, the insula metabolism-based treatment-specific biomarker defined in this study provides the first objective marker, to our knowledge, to guide initial treatment selection for depression.

Trial Registration: Registered at clinicaltrials.gov (NCT00367341)

JAMA Psychiatry.
Published online June 12, 2013.
doi:10.1001/jamapsychiatry.2013.143

MAJOR DEPRESSIVE DISORDER (MDD) is a highly prevalent, disabling, and costly illness.[1-3] For a patient presenting with MDD, an antidepressant medication or evidence-based psychotherapy is currently recommended as first-line treatment.[4-7] However, fewer than 40% of patients achieve remission with initial treatment,[8,9] and choosing the "wrong" initial treatment has significant individual and societal costs due to continued distress, risk of suicide, loss of productivity, and wasted resources associated with 2 to 3 months of an ineffective treatment.[10,11] Given the public health consequences of inadequately treated depression, a clinical or biological marker to guide initial treatment selection for MDD could have major health and economic impact.[12]

In other areas of medicine, identification of markers to guide treatment has significantly improved clinical outcome. For

Author Affiliations are listed at the end of this article.

©2013 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ by a Universita Torino User on 06/14/2013

example, in cancer[13] and heart disease,[14] biomarkers are currently used to optimize initial treatment selection as well as guide treatment modifications with disease progression. Over the past several decades, a number of potential markers to guide antidepressant treatment have been investigated including clinical,[15] immune,[16] inflammatory,[17] endocrine,[18] genetic,[19-21] and imaging/electroencephalography[22-27] measures. Despite this extensive research, to our knowledge, no clinically useful marker to guide treatment selection has emerged.

In the process of developing a marker to guide antidepressant treatment selection, it is important to consider what qualities such a marker should have. Toward this goal, a nonspecific biomarker that predicts improvement regardless of treatment is not useful. Rather, a clinically meaningful and treatment-specific biomarker (TSB) should (1) predict an individual's improvement to a specific treatment and (2) predict nonresponse to an alternative treatment. Such a biomarker can only be identified in a study that assesses outcome to 2 or more different treatments.

Previous neuroimaging studies have suggested that pretreatment brain activity patterns can predict efficacy, but those studies have generally focused on a particular treatment.[26,27] For example, higher rostral cingulate and/or subgenual cingulate activity has been associated with greater improvement with antidepressant medications,[28,29] sleep deprivation,[30] and cingulotomy.[31] Comparisons of different treatments have thus far identified markers of response and nonresponse but not patterns that differentiate among the treatments tested.[22,32,33] Further, imaging studies demonstrate that medications and psychotherapy have differential effects on distinct brain regions,[23,34] suggesting that baseline activity may indicate response to one treatment vs the other. Although, to our knowledge, no prior imaging study has directly assessed the association of pretreatment brain activity patterns with differential response to different treatments (eg, medication vs psychotherapy), these past studies strongly suggest that a neuroimaging-based TSB can be defined.

In this study, we measured pretreatment brain glucose metabolism in patients with MDD randomized to receive a selective serotonin reuptake inhibitor (escitalopram oxalate) or cognitive behavior therapy (CBT).[35,36] Positron emission tomography (PET) scan measurement of glucose metabolism was selected based on its high reliability and availability combined with its established use for studies of baseline scan patterns in depression and effects of various antidepressant treatments.[22,23,28,34,37-43] Our aim was to define an imaging TSB for these 2 potential first-line treatments, ie, a brain activity pattern that distinguishes escitalopram remitters from both escitalopram nonresponders and CBT remitters while concurrently distinguishing CBT remitters from both CBT nonresponders and escitalopram remitters.

METHODS

PATIENT SELECTION

Written informed consent was obtained from all participants, with the protocol conducted as approved by the Emory institutional review board and registered at clinicaltrials.gov (NCT00367341). Eligible participants were adult outpatients with a primary diagnosis of MDD as assessed by the Structured Clinical Interview for *DSM-IV-TR* Axis I Disorders, Research Version, Patient Edition With Psychotic Screen[44] and confirmed through a psychiatric evaluation conducted by a study psychiatrist. Patients aged 18 to 60 years were recruited through the Mood and Anxiety Disorders Program at Emory University via advertisements and clinician referrals.[45] Patients were required to have moderate to severe symptoms of depression, defined as a 17-item Hamilton Depression Rating Scale (HDRS)[46] score of 18 or more at screening and 15 or more at the baseline randomization visit. Exclusion criteria included a current diagnosis of a primary psychiatric disorder other than MDD; a medical or neurological condition that could contribute to depression or that might interfere with response to treatment such as chronic pain syndromes and irritable bowel syndrome; current suicidal ideation requiring urgent clinical intervention; comorbid substance abuse within the past 3 months; substance dependence within 12 months prior to the screening visit; current or intended pregnancy or breastfeeding; use of antidepressants within 7 days of the screening visit (5 weeks for fluoxetine); current psychotherapy at the time of screening; or receipt of electroconvulsive therapy within 6 months of the screening visit. Patients were also excluded if they had a lifetime history of failure to respond to 6 or more weeks of treatment with escitalopram oxalate (≥10 mg/d) or 4 or more sessions of CBT for depression.

TREATMENT PROTOCOL

Treatment consisted of 2 phases: a short-term treatment phase (phase 1) and a combination treatment phase (phase 2). Phase 1 provided the data for this report. In phase 1, patients were randomly assigned (1:1 ratio) to receive a 12-week treatment course of either escitalopram oxalate (flexibly dosed from 10-20 mg/d) or manual-based, depression-focused CBT (16 one-hour sessions over 12 weeks) (**Figure 1**). Prior to study start, the study statistician prepared a permuted-block randomization schedule, with the assignments placed in order and sealed in opaque envelopes. Following acquisition of the pretreatment PET and magnetic resonance imaging scans, patients who continued to meet eligibility criteria were randomized to escitalopram or CBT. Escitalopram oxalate was started at 10 mg/d and could be increased to 20 mg/d at or after week 3 if the patient had an HDRS score more than 7 and was tolerating the medication. Down-titration to 10 mg/d was permitted if adverse effects were intolerable at the 20–mg/d dose. The CBT sessions were scheduled twice weekly for the first 4 weeks, followed by weekly sessions for the subsequent 8 weeks. Changes in symptom severity were assessed using the HDRS conducted by raters blinded to treatment group. Ratings were performed weekly for the first 6 weeks and then biweekly until week 12. Patients who did not remit on completion of their phase 1 treatment were offered enrollment in phase 2 comprising an additional 12 weeks of treatment with combination escitalopram and CBT.

OUTCOME METRICS

Clinical outcomes were defined using the HDRS, with the target end point being remission, defined as an HDRS score of 7 or less at both weeks 10 and 12 of phase 1 treatment,[47] to ensure stability of remission beyond a single "good week." Nonresponse was defined as a 30% or less HDRS score change from baseline to the phase 1 end point.[48] Partial responders (change in HDRS score >30% but not achieving remission) and dropouts were not included in the analyses for this report to avoid

©2013 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ by a Universita Torino User on 06/14/2013

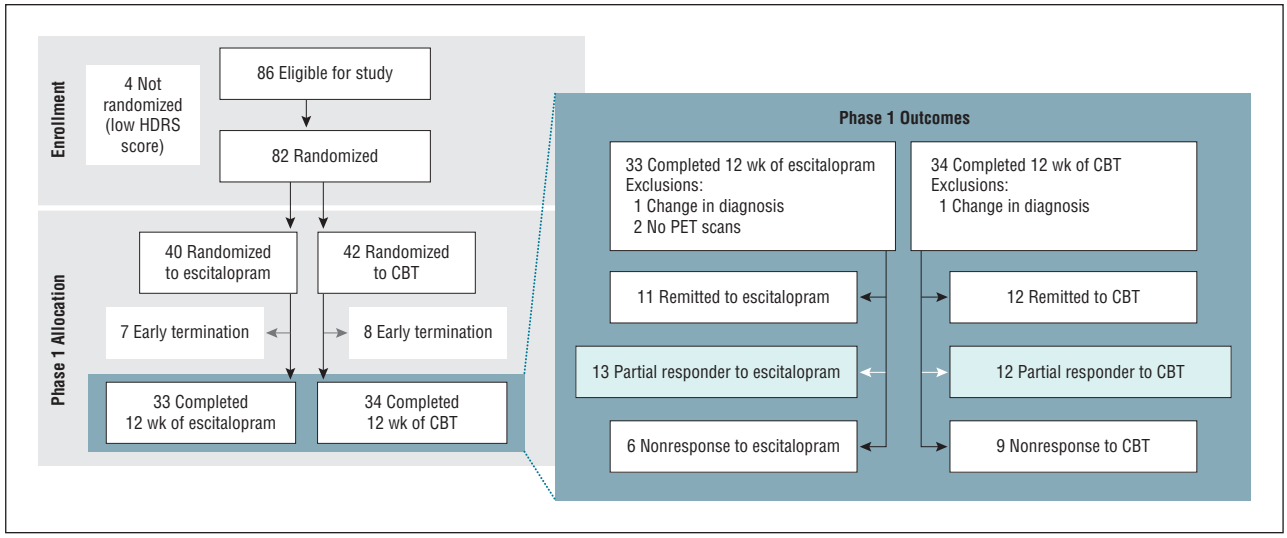


Figure 1. Study design and outcomes. Outcome groups defined by Hamilton Depression Rating Scale (HDRS) scores. Remission was defined as an HDRS score of 7 or less; partial response, as an HDRS score decrease of more than 30% but not achieving remission; and nonresponse, as an HDRS score decrease of 30% or less. Escitalopram was given as escitalopram oxalate. CBT indicates cognitive behavior therapy and PET, positron emission tomography.

potential dilution of either the remission or the nonresponse groups.

IMAGING ACQUISITION

Prior to treatment randomization, brain glucose metabolism was measured using PET (High-Resolution Research Tomograph scanner; Siemens), using standard methods without arterial blood sampling.[49] For each scan, a 370-MBq dose of fludeoxyglucose F18 (FDG) was administered intravenously, with a 20-minute 3-dimensional image acquisition beginning 40 minutes after tracer injection. During uptake, patients remained supine, awake, and resting with eyes closed and ears uncovered. Patients were given no explicit cognitive instructions but were asked to avoid ruminating on any 1 topic during the 40-minute FDG uptake period.[21] Raw emission images were corrected for injected dose and attenuation (using cesium 137, 6-minute transmission scan), reconstructed, and smoothed to an in-plane resolution of 4.0 mm Full-Width Half-Maximum.[34] A high-resolution T1-weighted structural magnetic resonance imaging scan was separately acquired for spatial normalization procedures and anatomical reference (TIM Trio 3-T whole-body scanner; Siemens) (3-dimensional magnetization-prepared rapid acquisition with gradient echo optimized at echo time = 5 milliseconds, repetition time = 35 milliseconds, matrix = $256 \times 208 \times 196$, and 1-mm isotropic resolution).

IMAGE PREPROCESSING

Attenuation-corrected PET images were coregistered to corresponding T1-weighted magnetic resonance imaging anatomical images using a 6-*df* linear transform and subsequently written into standard space using a nonlinear transform calculated from the T1-weighted image (DARTEL[50] and SPM8; Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/). Four patients had no anatomical scan and were normalized using a study-specific FDG template. Spatially normalized images were smoothed with an 8-mm Full-Width Half-Maximum Gaussian kernel and corrected for differences in whole-brain global mean activity.[23] Relative glucose metabolic rates were used for all analyses.

IMAGE ANALYSIS

A 2-way analysis of variance (ANOVA) with treatment (escitalopram or CBT) and outcome (remission or nonresponse) was performed to identify a putative escitalopram/CBT remission TSB using the baseline pretreatment FDG-PET scans (analyses performed with AFNI [National Institute of Mental Health, National Institutes of Health] and SPSS [IBM SPSS], statistical threshold $P < .001$ uncorrected, and a minimum cluster volume of 100 voxels, 0.34 mL). With this approach, a main effect of remission would identify brain regions associated with remission to treatment independent of randomization group, ie, a nonspecific biomarker. The treatment × outcome interaction would identify brain regions where the CBT treatment effect (remission or nonresponse) was distinguished from the escitalopram treatment effect (remission or nonresponse). Average normalized glucose metabolism values were extracted from clusters identified by the ANOVA (mean cluster activity) for further analysis.

Post hoc analyses of the extracted regions from the ANOVA interaction were used to refine selection of a potential TSB pattern by examining the effect sizes of the group differences for each region. We defined a region as a true TSB if it differentiated both the remission and nonresponse differences (by treatment) and the escitalopram and CBT differences (by outcome); thus, there were 4 comparisons of interest to consider when evaluating each region of interest as a stratification tool for treatment recommendation. Given the limited sample size, we report these comparisons using effect size, rather than statistical significance, to quantify their actual potential use as an eventual TSB. The 2-group effect size can be interpreted as the difference in metabolic activity between specified groups in units of standard deviation.[51] Because each region had a different magnitude of glucose metabolic activity and variation, each individual value was standardized using a *z* score, with regional *z* score means plotted to illustrate the nature of the regional interaction effects. Because these data are already standardized to the level of variation, there are no "error bars" in the related graphs.

To further assess the generalizability of findings identified in this restricted analysis to the full sample of study completers, metabolic activity was correlated with percentage of change in HDRS score within each treatment group to determine if the putative biomarkers identified in the ANOVA showed

©2013 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ by a Universita Torino User on 06/14/2013

Table 1. Group Comparisons of Clinical Characteristics

Variable	CBT		Escitalopram Oxalate		Significance of Group × Treatment Interaction
	Nonresponders (n = 9)	Remitters (n = 12)	Nonresponders (n = 6)	Remitters (n = 12)[a]	
Age, y, mean (SD)	45.4 (8.8)	42.5 (10.8)	40.3 (5.2)	39.8 (6.3)	.67[b]
Male, No. (%)	3 (33.3)	7 (58.3)	2 (33.3)	5 (41.7)	0.628[c]
White, No. (%)	7 (77.8)	8 (66.7)	5 (100)	8 (72.7)	0.345[c]
Education, y, mean (SD)	14.7 (1.7)	15.6 (1.6)	15.4 (1.7)	16.2 (2.1)	.94[b]
Age at onset of MDD, y, mean (SD)	28.4 (12.3)	28.7 (11.2)	24.0 (11.6)	25.5 (10.7)	.88[b]
≥3 Lifetime episodes, No. (%)	4 (50.0)	4 (33.3)	2 (33.3)	4 (33.3)	0.623[c]
Duration of current episode, wk, mean (SD)	124.5 (118.5)	257.3 (308.8)	299.8 (620.0)	135.3 (266.2)	.20[b]
No. of previous AD trials in current episode, mean (SD)	1.2 (1.5)	1.2 (1.0)	1.5 (1.4)	1.0 (1.1)	.59[b]
Melancholic subtype, No. (%)	4 (44.4)	5 (45.5)	3 (50.0)	6 (50.0)	0.976[c]
Current anxiety disorder, No. (%)	3 (33.3)	2 (16.7)	1 (16.7)	5 (41.7)	
Baseline HDRS score, mean (SD)	19.9 (3.8)	17.9 (2.7)	18.0 (2.1)	19.3 (3.7)	.13[b]
Baseline HAMA total score, mean (SD)	18.9 (7.6)	12.8 (2.8)	13.3 (2.3)	15.3 (3.0)	.009[b]
BDI total score, mean (SD)	19.2 (4.6)	18.7 (7.2)	20.0 (3.8)	21.5 (7.7)	.64[b]

Abbreviations: AD, antidepressant; BDI, Beck Depression Inventory; CBT, cognitive behavior therapy; HAMA, Hamilton Anxiety Rating Scale; HDRS, 17-item Hamilton Depression Rating Scale; MDD, major depressive disorder.

[a] Only 11 of 12 escitalopram remitters had usable positron emission tomography scans.

[b] *P* values from 2-way analysis of variance.

[c] *P* values from test of the homogeneity of the odds ratio (Breslow-Day).

the predicted general pattern in the full cohort of phase 1 treatment completers.

RESULTS

CLINICAL EFFECTS

Eighty-two patients were randomized to treatment; however, 2 patients received a change in their psychiatric diagnosis during the trial, and they were not used in the analyses. This resulted in 41 randomized to CBT and 39 to escitalopram. Sixty-five patients completed phase 1; 63 of the completers (79% of the total sample) had baseline FDG-PET scans available for analysis. Phase 1 remission rates were similar for both treatments: CBT = 12 of 33 (36.3%) and escitalopram = 12 of 30 (40.0%) (Figure 1 and **Table 1**). Nonresponse rates were also similar for both treatments: CBT = 9 of 33 (27.3%) and escitalopram = 6 of 30 (20.0%). Thirty-eight patients with clear outcomes and usable PET scans were included in the primary analysis: 12 patients with CBT remission, 11 patients with escitalopram remission, 9 patients with CBT nonresponse, and 6 patients with escitalopram nonresponse. There were no statistical differences in age, sex, or demographic or illness characteristics between randomization groups (escitalopram vs CBT). There were also no baseline demographic differences among the treatment-specific phase 1 outcome groups (Table 1). However, CBT nonresponders had higher baseline anxiety ratings (Hamilton Anxiety Rating Scale total score).

NEUROIMAGING RESULTS

Treatment × Outcome ANOVA

There was no significant main effect of remission, ie, no treatment-nonspecific biomarker was identified. Significant treatment × outcome interaction effects were demonstrated for 6 regions: right anterior insula, right inferior temporal cortex (Brodmann area [BA] 20), left amygdala, left premotor cortex (BA 6), right motor cortex (BA 4), and precuneus (BA 7) (**Table 2** and **Figure 2**).

Post hoc Analyses of Extracted Regions of Interest

The average effect sizes of each region for the various contrasts are shown in Table 2 in order of cluster size from the ANOVA. This was used to rank the regions of interest in the order of their potential utility as a discriminator. Only the insula and precuneus showed differences larger than 1 SD in all 4 contrasts, with the insula showing the largest average difference across all 4 comparisons. These findings indicate that metabolic activity of the right anterior insula is the most viable TSB candidate (Table 2 and **Figure 3**). Further, the anterior insula was the only region that showed relative hypometabolism in 1 group (region/whole-brain mean <1.0) and hypermetabolism in the other (region/whole-brain mean >1.0), adding support for potential use as a treatment stratification tool.

Assessment of the Insula TSB Across the Full Sample

There was a significant correlation between baseline insula activity and percentage of change in HDRS scores in both the CBT and escitalopram groups. A positive correlation was shown for the CBT group ($r = 0.55$; $df = 31$; $P = .001$) (Figure 3). In contrast, the escitalopram-treated patients showed an opposite but less significant correlation ($r = -0.31$; $df = 28$; $P = .09$). Both correlations are consistent with the more restricted findings in the binarized remitter-nonresponder analyses.

©2013 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ by a Universita Torino User on 06/14/2013

Table 2. Treatment by Outcome Interaction Results and Post hoc Analyses of Extracted Regions of Interest

Region	MNI Coordinates, Peak X	Y	Z	Side	Cluster Size,[a] Voxels	Effect Sizes[b] REM−NR to CBT	REM−NR to sCIT	CBT−sCIT in NR	CBT−sCIT in REM	Average Marginal ES
Anterior insula	+30.0	+24.0	−13.5	R	529	1.69	1.17	1.52	1.33	1.43
Inferior temporal (BA 20)	+42.0	−33.0	−25.5	R	469	1.23	1.45	2.09	0.59	1.34
Amygdala	−27.0	−7.5	−27.0	L	233	0.98	1.61	1.89	0.69	1.29
Premotor (BA 6)	−27.0	+1.5	+58.5	L	233	1.40	1.03	1.75	0.68	1.22
Motor (BA 4)	+25.5	−27.0	+60.0	R	147	0.61	1.78	1.80	0.58	1.19
Precuneus (BA 7)	−18.0	−67.5	+43.5	L	101	1.18	1.27	1.37	1.08	1.23

Abbreviations: BA, Brodmann area; CBT, cognitive behavior therapy; CBT−sCIT, mean difference between CBT and escitalopram oxalate (treatment effect); MNI, Montreal Neurological Institute; NR, nonresponders; REM, remitters; REM−NR, mean difference between remitters and nonresponders (remission effect); sCIT: escitalopram oxalate.

[a] Whole-brain 2-way analysis of variance with $P < .001$ uncorrected; voxel size, 1.5 mm × 1.5 mm × 1.5 mm.

[b] Effect size = mean difference divided by pooled standard deviation.

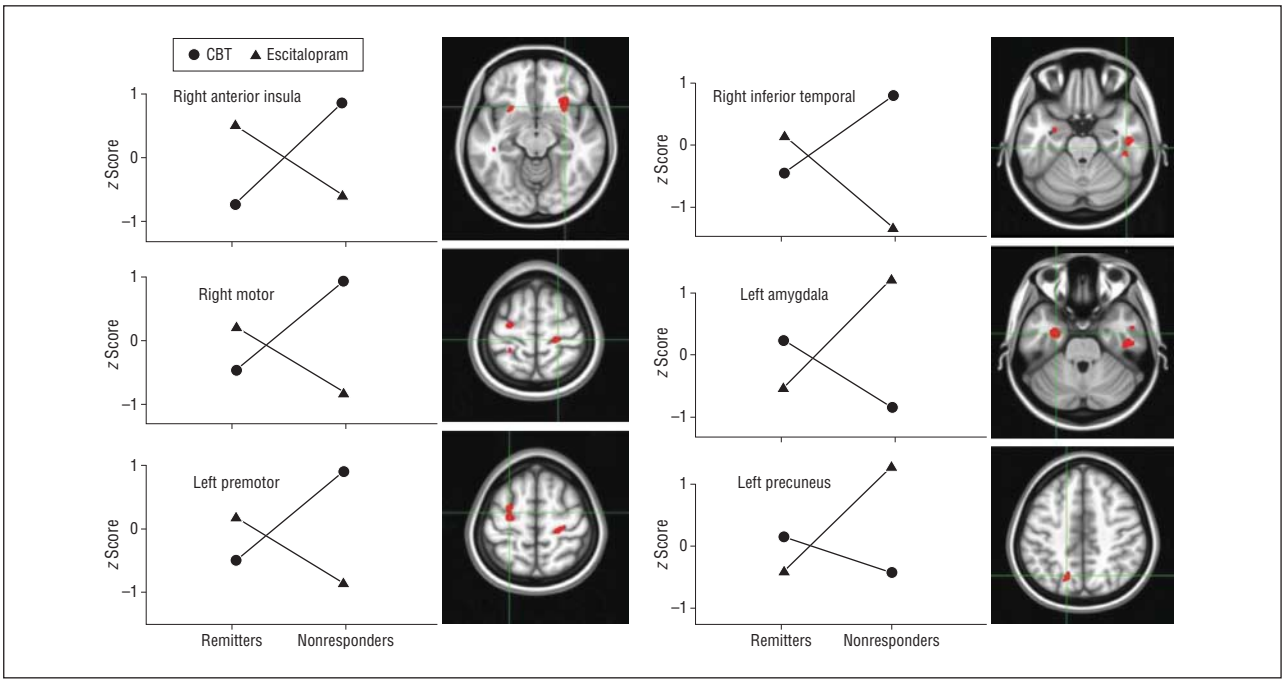


Figure 2. Potential treatment-specific biomarker candidates. Mean regional activity values for remitters and nonresponders segregated by treatment arm are plotted for the 6 regions showing a significant treatment × outcome analysis of variance interaction effect. Regional metabolic activity values are displayed as region/whole-brain metabolism converted to *z* scores. Regions match those shown in Table 2. Escitalopram was given as escitalopram oxalate. CBT indicates cognitive behavior therapy.

Although not a primary planned analysis, the presence of multiple regions identified in the ANOVA suggests that a combination rather than a single TSB might be more accurate in discriminating the groups. Although underpowered, a principal component analysis was performed using the 6 identified regions of interest. All regions loaded on 1 factor, which did not provide superior internal consistency to the insula alone (data not shown).

DISCUSSION

This 12-week randomized study of 2 first-line treatments for MDD identified 2 FDG-PET–defined brain pattern subtypes that differentially predicted remission to either CBT or escitalopram. Among the 6 identified cortical and limbic regions, the anterior insula metabolism best discriminated treatment outcome: insula hypometabolism was associated with remission to CBT and poor response to escitalopram, while insula hypermetabolism was associated with remission to escitalopram and poor response to CBT. These data suggest that insula metabolism alone (relative to each person's whole-brain mean metabolism) may serve as a pretreatment biomarker to guide initial treatment selection (medication vs CBT) for a patient presenting with a major depressive episode. To validate the insula TSB, a prospective replication study in which patients are treated according to brain type will be required. That said, this forced-choice analytic strategy establishes a potential stratification algorithm for man-

©2013 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ by a Universita Torino User on 06/14/2013

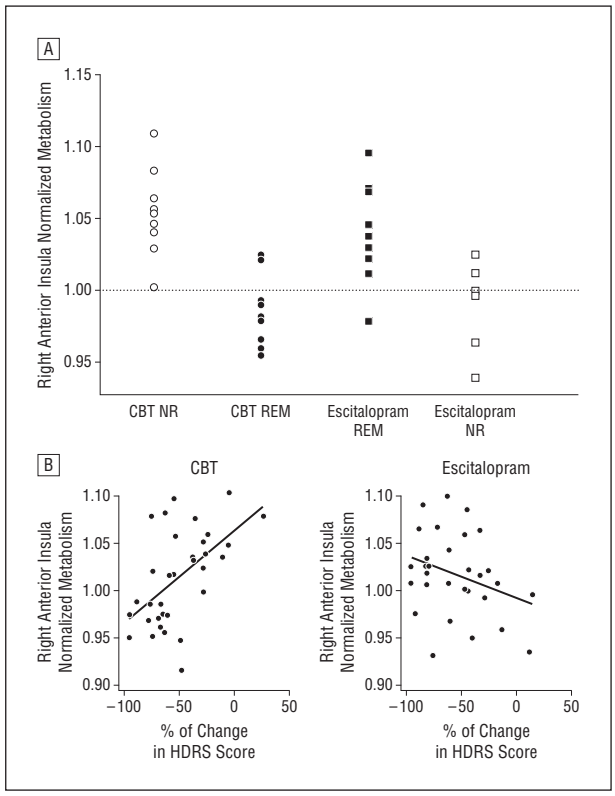


Figure 3. Right anterior insula as the optimal treatment-specific biomarker candidate. Expanded view of findings. A, Scatterplot of insular activity from individual subjects in the remitter (REM) and nonresponder (NR) groups. Note: the anterior insula is the only region where the interaction subdivides patients into hypermetabolic (region/whole-brain mean >1.0) and hypometabolic (region/whole-brain mean <1.0) subgroups. B, Correlations of insula activity with percentage of change in Hamilton Depression Rating Scale (HDRS) score in the full cohort of subjects treated with cognitive behavior therapy (CBT) and escitalopram oxalate.

aging patients with MDD based on brain state rather than patient or professional preference, anticipating the real-world decision-making process faced by clinicians, namely, choosing a first treatment that will most likely lead to remission while also avoiding a treatment that is likely to fail.

A role for the anterior insula in major depression is well established. The insula is crucial in mediating the translation of visceral experiences to subjective feeling states.[52] Additionally, anterior insula activity is linked to behaviors relevant to depression including interoception, emotional self-awareness, decision making, and cognitive control.[53-55] The anterior insula is extensively connected to various frontal, limbic, and brainstem regions, including the anterior cingulate cortex, amygdala, and hypothalamus.[56] Volume reductions of the anterior but not posterior insula have been described in currently ill patients with MDD as well as patients with remitted MDD compared with healthy controls.[57] Changes in insula activity occur with a variety of treatments for MDD, including medication,[58] vagus nerve stimulation,[59] deep brain stimulation,[60] and mindfulness training,[61] suggesting a role for this region in mediating antidepressant response and remission more generally.[62] Notably, past studies have reported both increases[33] and decreases[39] in baseline resting-state activity relative to never-depressed control subjects. This is consistent with the presence of at least 2 baseline patterns within the broader population of depressed patients. Most recently, baseline insula activity has been correlated with response to vagus nerve stimulation.[26] These previous studies taken together with the current findings support the anterior insula as a potential candidate for an imaging TSB.

Contrary to past published studies,[63] the rostral anterior cingulate did not discriminate the outcome subgroups in either the main effect or interaction analyses. A post hoc examination of responder and nonresponder differences within each treatment arm did reveal a nonsignificant rostral cingulate activity difference, with metabolism in responders greater than nonresponders, but solely in the escitalopram group. While consistent with past reports, this finding did not meet the TSB criteria defined for the current study, ie, a region whose activity can differentiate both good and poor outcomes for both treatments.

Critical to the stated aims, remission (rather than response) was the targeted end point in this study because the presence of residual symptoms is a known predictor of clinical relapse, even in patients with significant improvement.[64,65] Because the primary aim of this study was to identify distinct brain patterns that optimally predict remission to each of 2 specific treatments, patients with unclear treatment outcomes were excluded from the primary analysis (ie, responders without remission and partial responders). This enriched sample allowed for detection of clear remission and nonresponse signals; as such, these analyses did not attempt to characterize the neurobiological variability of patients with more ambiguous clinical outcomes. This is a commonly used approach when the goal is to develop or test a biological signal for stratifying subjects.[66,67] Nevertheless, baseline insula activity did correlate significantly with change in depression severity across all subjects, supporting the interpretation that insula activity is a plausible TSB suitable for further testing. Based on the correlational analysis across all subjects, the data further suggest that the anterior insula TSB may most optimally identify those patients who require CBT.

If confirmed with prospective testing, this putative TSB has both clinical and pathophysiological implications. At present, treatment failure with antidepressant medication often leads to the addition of a second drug and not a categorical switch to an evidence-based psychotherapy.[68] Results herein suggest that patients who require CBT have a distinct neurophysiology that differs categorically from patients who require escitalopram and knowledge of such may help to improve current clinical practice patterns. Further, using this or any other imaging-based TSB to define patient subgroups provides a brain-based platform to investigate genetic, immune, neuroendocrine, and behavioral variations from a new perspective.

While these first results are encouraging, there are several limitations. Clearly, there are patients who are not successfully treated with either of these 2 options, either alone or in combination.[69] Therefore, our strategy can be best seen as a first-line stratification approach to treat-

©2013 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ by a Universita Torino User on 06/14/2013

ment selection. Future studies, in addition to testing this insula biomarker prospectively, should include a design that works to identify patients resistant to both of these first-line treatments.[60,70]

The lack of a placebo arm could be considered a limitation, but given the randomized design of the study, there is no reason to believe that placebo responders would be unevenly distributed between the 2 groups. Thus, even if present, placebo effects on remission rates would be expected to be similar in both treatment groups. Although inclusion of a placebo arm might have provided further insights into mediators of improvement during treatment, the absence of a placebo arm does not diminish the potential clinical utility of the identified TSB.

It is also possible that these results are specific to the cohort recruited for this trial. As such, a stratification strategy based on insula metabolism will require prospective testing in a new group of comparably depressed patients. Similarly, additional studies will be required to determine if remitters to other medications have a similar or different TSB from that seen with escitalopram or if remitters to other evidence-based psychotherapies have a similar TSB to that seen with CBT.[71,72] Such studies are critical next steps toward the development of biology-based algorithms to guide treatment selection for MDD at all stages of illness. Still, if replicated, the insula TSB defined in this study would provide the first objective marker to guide initial treatment selection for major depression, an important advance in potentially reducing the costs and disability associated with this highly prevalent disorder.

Submitted for Publication: August 29, 2012; final revision received January 11, 2013; accepted January 14, 2013.
Published Online: June 12, 2013. doi:10.1001/jamapsychiatry.2013.143
Author Affiliations: Department of Psychiatry and Behavioral Sciences (Ms McGrath and Drs Dunlop, Craighead, and Mayberg), Graduate Division of Biological and Biomedical Sciences (Ms McGrath), Department of Biostatistics and Bioinformatics (Dr Kelley), and Department of Neurology (Dr Mayberg), Emory University, Atlanta, Georgia; Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire (Dr Holtzheimer); Department of Electrical Engineering and Brain Institute, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil (Dr Franco); and Center for the Developing Brain, Child Mind Institute, New York (Dr Craddock), and Nathan Kline Institute for Psychiatric Research, Orangeburg (Dr Craddock), New York.
Correspondence: Helen S. Mayberg, MD, Emory University, Department of Psychiatry, 101 Woodruff Cir, WMB 4313, Atlanta, GA 30322 (hmayber@emory.edu).
Author Contributions: Dr Mayberg had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Conflict of Interest Disclosures: Dr Dunlop has received honoraria for consulting work performed for Bristol-Myers Squibb, MedAvante, Pfizer, and Roche. He has received research support from AstraZeneca, Bristol-Myers Squibb, Evotec, Forest, GlaxoSmithKline, Novartis, Ono Pharmaceuticals, Pfizer, Shire, Takeda, and Transcept. Dr Holtzheimer has received consulting fees from St Jude Medical Neuromodulation and Cervel Neurotech and an honorarium from Johnson & Johnson. Dr Craighead is a board member of Hugarheill, an Icelandic company dedicated to the prevention of depression, and he receives book royalties from John Wiley & Sons. He is a consultant to the George West Mental Health Foundation that oversees Skyland Trail, a residential treatment facility in Atlanta, Georgia. Dr Mayberg has received consulting and intellectual property licensing fees from St Jude Medical Inc.
Funding/Support: This work was supported by National Institutes of Health grants R01 MH073719 (Dr Mayberg), T32 GM08695 (Ms McGrath), K23 MH086690 (Dr Dunlop), and K23 MH077869 (Dr Holtzheimer).
Additional Contributions: Treating psychotherapists: Sheethal Reddy, PhD, Patrick Sylvers, PhD, Lorie Ritschel, PhD, Meredith Jones, PhD, Mary Heekin, LCSW, Maryrose Gerardi, PhD, and Jill Rosenberg, LCSW. Treating physicians: Ebrahim Haroon, MD, Jeffrey Rakofsky, MD, Dylan Wint, MD, and Corey Beck, MD. Clinical coordinators: Ronald Chismar, RN, Melanie Galanti, BA, Rachelle Gibson, RN, Lauren Marx, BS, Melissa McKenzie, BA, and Tanja Mletzko, MA. Imaging team: Rebecca DeMayo, MA, Eundria Hill, MSW, Kiseung Choi, MS, and Justin Rajendra, BA. Raters: Margo Aaron, BA, Yara Betancourt, BA, Cristina Velasquez Delgado, BA, Novall Khan, BS, Ximena Marincic, BS, and Christopher Vaughan, BA.

REFERENCES

1. Kessler RC, Berglund P, Demler O, et al; National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*. 2003;289(23):3095-3105.
2. Birnbaum HG, Kessler RC, Kelley D, Ben-Hamadi R, Joish VN, Greenberg PE. Employer burden of mild, moderate, and severe major depressive disorder: mental health services utilization and costs, and work performance. *Depress Anxiety*. 2010;27(1):78-89.
3. Judd LL, Akiskal HS, Zeller PJ, et al. Psychosocial disability during the long-term course of unipolar major depressive disorder. *Arch Gen Psychiatry*. 2000; 57(4):375-380.
4. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed, text revision. Washington, DC: American Psychiatric Association; 2000.
5. Kennedy SH, Lam RW, Parikh SV, Patten SB, Ravindran AV. Canadian Network for Mood and Anxiety Treatments (CANMAT) clinical guidelines for the management of major depressive disorder in adults: introduction. *J Affect Disord*. 2009; 117(suppl 1):S1-S2.
6. American Psychiatric Association. *Treating Major Depressive Disorder: A Quick Reference Guide*. Washington, DC: American Psychiatric Association; 2010:1-28.
7. National Institute for Health and Clinical Excellence. *Depression: The Treatment and Management of Depression in Adults*. London, England: National Institute for Health and Clinical Excellence; 2009. NICE clinical guideline 90.
8. Gaynes BN, Warden D, Trivedi MH, Wisniewski SR, Fava M, Rush AJ. What did STAR*D teach us? results from a large-scale, practical, clinical trial for patients with depression. *Psychiatr Serv*. 2009;60(11):1439-1445.
9. Holtzheimer PE, Mayberg HS. Stuck in a rut: rethinking depression and its treatment. *Trends Neurosci*. 2011;34(1):1-9.
10. Kessler RC, Akiskal HS, Ames M, et al. Prevalence and effects of mood disorders on work performance in a nationally representative sample of U.S. workers. *Am J Psychiatry*. 2006;163(9):1561-1568.
11. Dunlop BW, Reddy S, Yang L, Lubaczewski S, Focht K, Guico-Pabia CJ. Symptomatic and functional improvement in employed depressed patients: a double-blind clinical trial of desvenlafaxine versus placebo. *J Clin Psychopharmacol*. 2011; 31(5):569-576.

©2013 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ by a Universita Torino User on 06/14/2013

12. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry*. 2012;17(12):1174-1179.
13. Saijo N. Critical comments for roles of biomarkers in the diagnosis and treatment of cancer. *Cancer Treat Rev*. 2012;38(1):63-67.
14. Welch TD, Yang EH, Reeder GS, Gersh BJ. Modern management of acute myocardial infarction. *Curr Probl Cardiol*. 2012;37(7):237-310.
15. Quitkin FM, Stewart JW, McGrath PJ, et al. Columbia atypical depression: a subgroup of depressives with better response to MAOI than to tricyclic antidepressants or placebo. *Br J Psychiatry Suppl*. 1993;(21):30-34.
16. Irwin MR, Miller AH. Depressive disorders and immunity: 20 years of progress and discovery. *Brain Behav Immun*. 2007;21(4):374-383.
17. Müller N, Myint A-M, Schwarz MJ. Inflammatory biomarkers and depression. *Neurotox Res*. 2011;19(2):308-318.
18. Arana GW, Baldessarini RJ, Ornsteen M. The dexamethasone suppression test for diagnosis and prognosis in psychiatry: commentary and review. *Arch Gen Psychiatry*. 1985;42(12):1193-1204.
19. Huezo-Diaz P, Uher R, Smith R, et al. Moderation of antidepressant response by the serotonin transporter gene. *Br J Psychiatry*. 2009;195(1):30-38.
20. Ising M, Lucae S, Binder EB, et al. A genomewide association study points to multiple loci that predict antidepressant drug treatment outcome in depression. *Arch Gen Psychiatry*. 2009;66(9):966-975.
21. D'Empaire I, Guico-Pabia CJ, Preskorn SH. Antidepressant treatment and altered CYP2D6 activity: are pharmacokinetic variations clinically relevant? *J Psychiatr Pract*. 2011;17(5):330-339.
22. Konarski JZ, Kennedy SH, Segal ZV, et al. Predictors of nonresponse to cognitive behavioural therapy or venlafaxine using glucose metabolism in major depressive disorder. *J Psychiatry Neurosci*. 2009;34(3):175-180.
23. Kennedy SH, Konarski JZ, Segal ZV, et al. Differences in brain glucose metabolism between responders to CBT and venlafaxine in a 16-week randomized controlled trial. *Am J Psychiatry*. 2007;164(5):778-788.
24. DeBattista C, Kinrys G, Hoffman D, et al. The use of referenced-EEG (rEEG) in assisting medication selection for the treatment of depression. *J Psychiatr Res*. 2011;45(1):64-75.
25. Leuchter AF, Cook IA, Marangell LB, et al. Comparative effectiveness of biomarkers and clinical indicators for predicting outcomes of SSRI treatment in Major Depressive Disorder: results of the BRITE-MD study. *Psychiatry Res*. 2009;169(2):124-131.
26. Conway CR, Chibnall JT, Gangwani S, et al. Pretreatment cerebral metabolic activity correlates with antidepressant efficacy of vagus nerve stimulation in treatment-resistant major depression: a potential marker for response? *J Affect Disord*. 2012;139(3):283-290.
27. Siegle GJ, Thompson WK, Collier A, et al. Toward clinically useful neuroimaging in depression treatment: prognostic utility of subgenual cingulate activity for determining depression outcome in cognitive therapy across studies, scanners, and patient characteristics. *Arch Gen Psychiatry*. 2012;69(9):913-924.
28. Mayberg HS, Brannan SK, Mahurin RK, et al. Cingulate function in depression: a potential predictor of treatment response. *Neuroreport*. 1997;8(4):1057-1061.
29. Pizzagalli D, Pascual-Marqui RD, Nitschke JB, et al. Anterior cingulate activity as a predictor of degree of treatment response in major depression: evidence from brain electrical tomography analysis. *Am J Psychiatry*. 2001;158(3):405-415.
30. Wu J, Buchsbaum MS, Gillin JC, et al. Prediction of antidepressant effects of sleep deprivation by metabolic rates in the ventral anterior cingulate and medial prefrontal cortex. *Am J Psychiatry*. 1999;156(8):1149-1158.
31. Dougherty DD, Weiss AP, Cosgrove GR, et al. Cerebral metabolic correlates as potential predictors of response to anterior cingulotomy for treatment of major depression. *J Neurosurg*. 2003;99(6):1010-1017.
32. Brody AL, Saxena S, Stoessel P, et al. Regional brain metabolic changes in patients with major depression treated with either paroxetine or interpersonal therapy: preliminary findings. *Arch Gen Psychiatry*. 2001;58(7):631-640.
33. Ketter TA, Kimbrell TA, George MS, et al. Baseline cerebral hypermetabolism associated with carbamazepine response, and hypometabolism with nimodipine response in mood disorders. *Biol Psychiatry*. 1999;46(10):1364-1374.
34. Goldapple K, Segal Z, Garson C, et al. Modulation of cortical-limbic pathways in major depression: treatment-specific effects of cognitive behavior therapy. *Arch Gen Psychiatry*. 2004;61(1):34-41.
35. Beck A, Rush A, Shaw B, Emery G. *Cognitive Therapy of Depression*. New York, NY: Guilford; 1979.
36. Beck AT. The current state of cognitive therapy: a 40-year retrospective. *Arch Gen Psychiatry*. 2005;62(9):953-959.
37. Bartlett EJ, Barouche F, Brodie JD, et al. Stability of resting deoxyglucose metabolic values in PET studies of schizophrenia. *Psychiatry Res*. 1991;40(1):11-20.
38. Brody AL, Saxena S, Silverman DH, et al. Brain metabolic changes in major depressive disorder from pre- to post-treatment with paroxetine. *Psychiatry Res*. 1999;91(3):127-139.
39. Kimbrell TA, Ketter TA, George MS, et al. Regional cerebral glucose utilization in patients with a range of severities of unipolar depression. *Biol Psychiatry*. 2002;51(3):237-252.
40. Drevets WC, Price JL, Bardgett ME, Reich T, Todd RD, Raichle ME. Glucose metabolism in the amygdala in depression: relationship to diagnostic subtype and plasma cortisol levels. *Pharmacol Biochem Behav*. 2002;71(3):431-447.
41. Saxena S, Brody AL, Ho ML, Zohrabi N, Maidment KM, Baxter LR Jr. Differential brain metabolic predictors of response to paroxetine in obsessive-compulsive disorder versus major depression. *Am J Psychiatry*. 2003;160(3):522-532.
42. Little JT, Ketter TA, Kimbrell TA, et al. Bupropion and venlafaxine responders differ in pretreatment regional cerebral metabolism in unipolar depression. *Biol Psychiatry*. 2005;57(3):220-228.
43. Milak MS, Parsey RV, Lee L, et al. Pretreatment regional brain glucose uptake in the midbrain on PET may predict remission from a major depressive episode after three months of treatment. *Psychiatry Res*. 2009;173(1):63-70.
44. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for* DSM-IV-TR *Axis I Disorders, Research Version, Patient Edition With Psychotic Screen (SCID-I/P W/PSY SCREEN)*. New York: Biometrics Research, New York State Psychiatric Institute; 2002.
45. Dunlop BW, Kelley ME, Mletzko TC, Velasquez CM, Craighead WE, Mayberg HS. Depression beliefs, treatment preference, and outcomes in a randomized trial for major depressive disorder. *J Psychiatr Res*. 2012;46(3):375-381.
46. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23:56-62.
47. Rush AJ, Trivedi MH, Stewart JW, et al. Combining Medications to Enhance Depression Outcomes (CO-MED): acute and long-term outcomes of a single-blind randomized study. *Am J Psychiatry*. 2011;168(7):689-701.
48. Nierenberg AA, Farabaugh AH, Alpert JE, et al. Timing of onset of antidepressant response with fluoxetine treatment. *Am J Psychiatry*. 2000;157(9):1423-1428.
49. Phelps ME, Huang SC, Hoffman EJ, Selin C, Sokoloff L, Kuhl DE. Tomographic measurement of local cerebral glucose metabolic rate in humans with (F-18)2-fluoro-2-deoxy-D-glucose: validation of method. *Ann Neurol*. 1979;6(5):371-388.
50. Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage*. 2007;38(1):95-113.
51. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associated; 1988.
52. Critchley HD, Wiens S, Rotshtein P, Ohman A, Dolan RJ. Neural systems supporting interoceptive awareness. *Nat Neurosci*. 2004;7(2):189-195.
53. Farb NA, Segal ZV, Anderson AK. Attentional modulation of primary interoceptive and exteroceptive cortices. *Cereb Cortex*. 2013;23(1):114-126.
54. Craig AD. How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci*. 2009;10(1):59-70.
55. Critchley HD. Neural mechanisms of autonomic, affective, and cognitive integration. *J Comp Neurol*. 2005;493(1):154-166.
56. Augustine JR. Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Res Brain Res Rev*. 1996;22(3):229-244.
57. Takahashi T, Yücel M, Lorenzetti V, et al. Volumetric MRI study of the insular cortex in individuals with current and past major depression. *J Affect Disord*. 2010;121(3):231-238.
58. Kennedy SH, Evans KR, Krüger S, et al. Changes in regional brain glucose metabolism measured with positron emission tomography after paroxetine treatment of major depression. *Am J Psychiatry*. 2001;158(6):899-905.
59. Conway CR, Sheline YI, Chibnall JT, George MS, Fletcher JW, Mintun MA. Cerebral blood flow changes during vagus nerve stimulation for depression. *Psychiatry Res*. 2006;146(2):179-184.
60. Mayberg HS, Lozano AM, Voon V, et al. Deep brain stimulation for treatment-resistant depression. *Neuron*. 2005;45(5):651-660.
61. Farb NAS, Segal ZV, Mayberg H, et al. Attending to the present: mindfulness meditation reveals distinct neural modes of self-reference. *Soc Cogn Affect Neurosci*. 2007;2(4):313-322.
62. Fu CH, Steiner H, Costafreda SG. Predictive neural biomarkers of clinical response in depression: a meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. *Neurobiol Dis*. 2013;52:75-83.
63. Pizzagalli DA. Frontocingulate dysfunction in depression: toward biomarkers of treatment response. *Neuropsychopharmacology*. 2011;36(1):183-206.
64. Paykel ES, Ramana R, Cooper Z, Hayhurst H, Kerr J, Barocka A. Residual symptoms after partial remission: an important outcome in depression. *Psychol Med*. 1995;25(6):1171-1180.
65. Judd LL, Paulus MJ, Schettler PJ, et al. Does incomplete recovery from first life-

©2013 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ by a Universita Torino User on 06/14/2013

time major depressive episode herald a chronic course of illness? *Am J Psychiatry*. 2000;157(9):1501-1504.
66. Ridker PM, Hennekens CH, Buring JE, Rifai N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *N Engl J Med*. 2000;342(12):836-843.
67. Ridker PM. Cardiology Patient Page. C-reactive protein: a simple test to help predict risk of heart attack and stroke. *Circulation*. 2003;108(12):e81-e85.
68. Gaynes BN, Dusetzina SB, Ellis AR, et al. Treating depression after initial treatment failure: directly comparing switch and augmenting strategies in STAR*D. *J Clin Psychopharmacol*. 2012;32(1):114-119.
69. Thase ME, Friedman ES, Biggs MM, et al. Cognitive therapy versus medication in augmentation and switch strategies as second-step treatments: a STAR*D report. *Am J Psychiatry*. 2007;164(5):739-752.
70. Rush AJ, Warden D, Wisniewski SR, et al. STAR*D: revising conventional wisdom. *CNS Drugs*. 2009;23(8):627-647.
71. Dunlop BW, Binder EB, Cubells JF, et al. Predictors of remission in depression to individual and combined treatments (PReDICT): study protocol for a randomized controlled trial. *Trials*. 2012;13(1):106.
72. Kennedy SH, Downar J, Evans KR, et al. The Canadian Biomarker Integration Network in Depression (CAN-BIND): advances in response prediction. *Curr Pharm Des*. 2012;18(36):5976-5989.

©2013 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ by a Universita Torino User on 06/14/2013

Neuron, Vol. 45, 651–660, March 3, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.neuron.2005.02.014

Deep Brain Stimulation for Treatment-Resistant Depression

Clinical Study

Helen S. Mayberg,[1,2,*] Andres M. Lozano,[3,*] Valerie Voon,[4] Heather E. McNeely,[5] David Seminowicz,[6] Clement Hamani,[3] Jason M. Schwalb,[3] and Sidney H. Kennedy[4]

[1]Rotman Research Institute at Baycrest Centre
and Departments of Psychiatry and Neurology
University of Toronto
Toronto, Ontario, M6A 2E1
Canada

[2]Emory University School of Medicine
Atlanta, Georgia 30322

[3]Division of Neurosurgery and
Toronto Western Hospital Research Institute
University Health Network
University of Toronto
Toronto, Ontario, M5P 2P2
Canada

[4]Department of Psychiatry
University Health Network
University of Toronto
Toronto, Ontario, M5P 2P2
Canada

[5]Neuropsychology Lab
Department of Psychiatry
Center for Addiction and Mental Health
University of Toronto
Toronto, Ontario, M5T 1R8
Canada

[6]Rotman Research Institute at Baycrest Centre
and the Institute of Medical Science
University of Toronto
Toronto, Ontario, M6A 2E1
Canada

Summary

Treatment-resistant depression is a severely disabling disorder with no proven treatment options once multiple medications, psychotherapy, and electroconvulsive therapy have failed. Based on our preliminary observation that the subgenual cingulate region (Brodmann area 25) is metabolically overactive in treatment-resistant depression, we studied whether the application of chronic deep brain stimulation to modulate BA25 could reduce this elevated activity and produce clinical benefit in six patients with refractory depression. Chronic stimulation of white matter tracts adjacent to the subgenual cingulate gyrus was associated with a striking and sustained remission of depression in four of six patients. Antidepressant effects were associated with a marked reduction in local cerebral blood flow as well as changes in downstream limbic and cortical sites, measured using positron emission tomography. These results suggest that disrupting focal pathological activity in limbic-cortical circuits using electrical stimulation of the subgenual cingulate white matter can effectively reverse symptoms in otherwise treatment-resistant depression.

*Correspondence: hmayber@emory.edu (H.S.M.); lozano@uhnres.utoronto.ca (A.M.L.)

Introduction

Major depression is the most common of all psychiatric disorders (Wang, 2003). It ranks among the top causes of worldwide disease burden and is the leading source of disability in adults in North America under the age of 50 (World Health Organization, 2001). While depression can be effectively treated in the majority of patients by either medication or some form of evidence-based psychotherapy (American Psychiatric Association, 2000), up to 20% of patients fail to respond to standard interventions (Fava, 2003; Keller et al., 1992). For these patients, trial-and-error combinations of multiple medications and electroconvulsive therapy are often required (Kennedy and Lam, 2003; Lancet UK ECT Review Group, 2003; Sackeim et al., 2001). For patients who remain severely depressed despite these aggressive approaches, new strategies are needed. We describe a strategy directed at this group of treatment-refractory patients with major depression.

Converging clinical, biochemical, neuroimaging, and postmortem evidence suggests that depression is unlikely to be a disease of a single brain region or neurotransmitter system. Rather, it is now generally viewed as a systems-level disorder affecting integrated pathways linking select cortical, subcortical, and limbic sites and their related neurotransmitter and molecular mediators (Manji et al., 2001; Mayberg, 1997; Nemeroff 2002; Nestler et al., 2002; Vaidya and Duman, 2001). While mechanisms driving this "system dysfunction" are not yet characterized, they are likely to be multifactorial, with genetic vulnerability, developmental insults, and environmental stressors all considered important and synergistic contributors (Caspi et al., 2003; Heim et al., 2000; Kendler et al., 2001). Treatments for depression can be similarly viewed within this limbic-cortical system framework, where different modes of treatment modulate specific regional targets, resulting in a variety of complementary, adaptive chemical and molecular changes that re-establish a normal mood state (Vaidya and Duman 2001; Hyman and Nestler, 1996, Mayberg 2003).

Functional neuroimaging studies have had a critical role in characterizing these limbic-cortical pathways (Brody, et al., 2001; Drevets, 1999; Mayberg, 1994, 2003). Our own studies have demonstrated consistent involvement of the subgenual cingulate (Cg25) in both acute sadness and antidepressant treatment effects, suggesting a critical role for this region in modulating negative mood states (Mayberg et al., 1999; Seminowicz et al., 2004). In support of this hypothesis, a decrease in Cg25 activity is reported with clinical response to different antidepressant treatments including specific serotonin reuptake inhibitor (SSRI) antidepressant medications, electroconvulsive therapy (ECT), re-

petitive transcranial magnetic stimulation (rTMS), and ablative surgery (Dougherty et al., 2003; Goldapple et al., 2004; Malizia, 1997; Mayberg et al., 2000; Mottaghy et al., 2002; Nobler et al., 2001).

In addition, Cg25 connections to the brainstem, hypothalamus, and insula have been implicated in the disturbances of circadian regulation associated with depression (sleep, appetite, libido, neuroendocrine changes) (Barbas et al., 2003; Freedman et al., 2000; Jurgens and Muller-Preuss, 1977; Maclean, 1990; Ongur et al., 1998). Reciprocal pathways linking Cg25 to orbitofrontal, medial prefrontal, and various parts of the anterior and posterior cingulate cortices form the neuroanatomical substrates by which primary autonomic and homeostatic processes influence various aspects of learning, memory, motivation and reward—core behaviors altered in depressed patients (Barbas et al., 2003; Carmichael and Price, 1996; Haber, 2003; Vogt and Pandya, 1987).

Recent advances in the surgical treatment of Parkinson's disease have demonstrated that chronic high-frequency deep brain stimulation (DBS) in pathologically overactive brain circuits produces profound clinical benefits (Benabid, 2003; Lang and Lozano, 1998). We have previously shown that clinically effective DBS in the basal ganglia produces both local and remote changes in neural activity as assessed by positron emission tomography (PET) (Davis et al., 1997). We have now leveraged these observations to investigate whether DBS could be used to modulate pathological brain circuits in depression.

We tested the hypothesis that the use of chronic stimulation to modulate Cg25 gray matter and interconnected frontal and subcortical regions could reverse the pathological metabolic activity in these circuits and produce clinical benefits in patients with treatment-resistant depression (TRD). This study reports the use of high-frequency subgenual cingulate white matter (Cg25WM) DBS in six TRD patients.

Results

Patients

All six patients met DSM IV-TR criteria for major depressive disorder (MDD) with a major depressive episode (MDE) of at least 1 year duration diagnosed by structured clinical interview for DSM IV-TR (First et al., 2001), and all had severe depression with a minimum score at entry of 20 on the 17 item Hamilton Depression Rating Scale (HDRS) (Hamilton, 1960). Each met stringent criteria for treatment resistance defined as failure to response to a minimum of four different antidepressant treatments, including medications, evidence-based psychotherapy, or electroconvulsive therapy, administered at adequate doses and duration during the current episode (Fava, 2003; Sackeim et al., 2001; Thase and Rush 1997). Subjects' demographic characteristics are presented in Table 1.

Intraoperative Findings

DBS electrodes were implanted in Cg25WM under local anesthesia using MR imaging guidance (Figure 1, row 1: A/B and row 2: C/D) (Schaltenbrand and Wahren, 1977). A protocol similar to that used in evaluating stimulation thresholds for efficacy or adverse effects in patients with Parkinson's disease was adopted for use in these patients (Benabid, 2003; Davis et al., 1997; Lang and Lozano, 1998). The spontaneous report or occurrence of any acute behavioral, cognitive, motor, or autonomic effects was sought during blinded, sequential stimulation of successive, individual contacts (monopolar stimulation, 60 μs pulsewidths, 130 Hz). Voltage was progressively increased up to 9.0 V at each of the eight electrode contacts (four per side), as tolerated. Voltage was increased by approximately 1.0 V every 30 s, with a 15–20 s pause between adjustments, allowing time for patients to identify an effect, if present. Patients reported no motor or sensory phenomena that cued them as to whether current was either on or off.

All patients spontaneously reported acute effects including "sudden calmness or lightness," "disappearance of the void," sense of heightened awareness, increased interest, "connectedness," and sudden brightening of the room, including a description of the sharpening of visual details and intensification of colors in response to electrical stimulation. Reproducible and reversible changes in these phenomena, time locked with stimulation, were observed at specific contacts and parameters for individual patients and not with sham or subthreshold stimulation at those same sites. Increases in motor speed, volume, and rate of spontaneous speech and improved prosody were observed. In addition, changes in both positive and negative affective rating scores (PANAS scale) (Watson and Clark, 1988) occurred coincident with the patients' spontaneous statements. There were no overt adverse affective or autonomic changes with stimulation at settings producing these improvements. However, all patients experienced stimulation dose-dependent adverse effects including lightheadedness and psychomotor slowing at high settings (over 7.0 Volts), most often seen at the superior electrode contact.

Postoperative Findings: Short-Term Stimulation Effects

Postoperative MR imaging confirmed the placement of the DBS electrodes within the subgenual cingulate white matter (Cg25WM) bilaterally as targeted. (Figure 1, row 3: E/F). During the 5 day postoperative period, and prior to placement of the pulse generator, daily short sessions of DBS were used to refine final contact selection and stimulation parameters. Systematic testing of individual and paired unilateral and bilateral contacts was performed with a variety of parameters (monopolar [contact anode; case cathode] and bipolar, pulsewidth of 30–250 μs, frequency of 10–130 Hz, progressive increase in voltage from 0.0–9.0 Volts). Acute behavioral changes were again observed during these test sessions. Reproducible improvements in interest, motor speed, activity level, and PANAS scores (reduced negative, increased positive scores) were seen during these stimulation sessions generally using the same contacts and parameters that induced effects in the operating room.

To control for the possibility of a placebo effect, patients were blinded as to which contact was being stim-

Table 1. Patient Demographics

Patient #	1	2	3	4	5	6	Group
Gender	F	M	F	M	M	F	3F/3M
Current age	48	59	45	48	37	39	46 ± 8
Age MDD onset	18	45	21	40	19	34	29.5 ± 12
Current episode (yrs)	1.5	3	6	8	10	5	5.6 ± 3
# Lifetime episodes	12	9	3	1	2	1	4.7 ± 5
Hamilton depression score (17 item)	29	20	27	24	26	25	25 ± 3
Past ECT	no	yes	yes	yes	yes	yes	5 of 6
Past psychotherapy	yes	yes	yes	yes	yes	yes	6 of 6
Family history MDD	yes	no	yes	yes	yes	yes	5 of 6
DSM IV diagnosis	UP	BPII	UP	UP	UP	UP	5 UP
Melancholic subtype	yes	no	yes	no	yes	yes	4 of 6
Current medications[a]	1–5	1, 3	1–4	1, 2, 4, 6, 7	1, 7	1, 4, 5, 7	

MDD, major depressive disorder; ECT, electroconvulsive therapy; DSM IV, Diagnostic and Statistical Manual of Mental Disorders, Version IV; UP, unipolar; BP, bipolar.
[a] Current medications: 1, SSRI/SNRI; 2, bupropion; 3, atypical antipsychotic; 4, benzodiazepine; 5, stimulant; 6, mood stabilizer; 7, other.

ulated and to the parameter settings. Sham stimulation using either 0.0 Volts or subthreshold stimulation failed to elicit any changes in behavior. Randomized, acute *off-on-off-on* trials of varying short periods of stimulation (1–5 min) at optimal settings produced consistent stimulation-locked behavioral improvements. Unexpectedly, with application of stimulation for progressively longer periods (from 1 to 3 hr), there was an increasing and correspondingly longer carry-over of the beneficial behavioral effects beyond cessation of the stimulation. These longer stimulation periods were free of adverse effects and were the basis for selecting the stimulation settings used chronically.

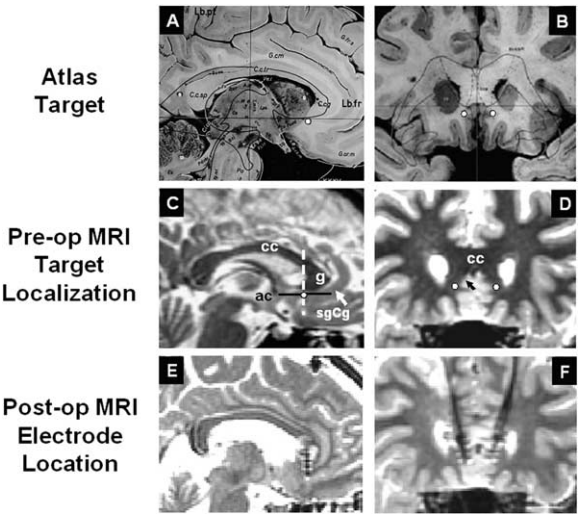


Figure 1. DBS Electrode Placement in the Subgenual Cingulate White Matter

(Row 1) Sagittal (left, [A]) and coronal (right, [B]) views of the subgenual cingulate target (white circles) localized on the Schaltenbrandt neurosurgical atlas. (Row 2) Sagittal (C) and coronal (D) views of the DBS target mapped on a high-resolution T1 MRI scans for one patient. (Row 3) Sagittal (E) and coronal (F) views of postop MRI scans demonstrating the location of electrodes for a single subject with the ventral contact centered within the predetermined location. sgCg, subgenual cingulate; cc, corpus callosum; g, genu of the corpus callosum; ac, anterior commissure; white circles, electrode target in sgCg white matter; white and black arrows, sgCg gyrus; dotted line, anterior-posterior position of the electrode relative to the ac-g line.

Postoperative Selection of Stimulation Parameters

Patients were discharged home with stimulation "off" following implantation of the pulse generators. One week later, chronic DBS was initiated using the lowest voltage and specific electrode contacts that had previously produced acute behavioral effects. Parameters of stimulation were reassessed at weekly intervals with minor adjustments in voltage made to optimize clinical effects. Following a 4 week period of parameter optimization, settings generally remained stable for the remainder of the 6 month follow-up period. The mean stimulation parameters used in this group at 6 months were 4.0 Volts, 60 μs pulsewidths, at a frequency of 130 Hz.

Efficacy of Chronic Bilateral Cg25WM DBS

Standard criteria for antidepressant response and remission were applied (Frank et al., 1991). Response was defined as a decrease in the HDRS-17 score of 50% or greater from the pretreatment baseline; remission as an absolute HDRS-17 score <8. One month postop, two patients met criteria for clinical response (Table 2). By 2 months, five of the six patients met the defined response threshold. Continued antidepressant response was seen in four of these subjects, with some variability up to 5 months. At the 6 month study endpoint, antidepressant response was maintained in four subjects (66%). Moreover, three of these subjects achieved remission or near remission of illness. Consistent with the improvements seen in the HDRS-17 scores, comparable changes were also demonstrated on other quantitative depression scales (see Table 3). Presurgical medications and doses were unchanged throughout the 6 month study.

Normalization of early-morning sleep disturbance (middle insomnia commonly seen in MDD) occurred in the first week of chronic DBS in four of the six subjects (patients 1, 3, 5, and 6) and was the first notable sus-

Table 2. Hamilton Depression Rating Scale, HDRS-17, Scores over Time for Each Subject

Time	Hamilton Score[a]					
	Pt 1[b]	Pt 2[c]	Pt 3[b]	Pt 4[c]	Pt 5[b]	Pt 6[b]
Preop baseline	29	22	29	24	26	25
1 week postop (acute stimulation)	5	10	12	18	17	12
2 weeks postop (DBS off)	9	13	23	18	22	n/a
1 month	10	14	17	20	22	12
2 months	13	11	12	18	10	12
3 months	2	15	14	25	7	14
4 months	4	9	12	24	6	12
5 months	5	18	7	23	8	n/a
6 months	5	15	9	23	6	12

[a] Clinical response: decrease HDRS score >50%. Clinical remission: absolute HDRS score <8.
[b] Clinical responders.
[c] Clinical nonresponders.

tained symptom change. Over the initial few weeks of continuous DBS, increased energy, interest, and psychomotor speed were additionally reported, with effects generally appreciable a day or two following stimulation adjustments. Patients and their families described renewed interest and pleasure in social and family activities, decreased apathy and anhedonia, as well as an improved ability to plan, initiate, and complete tasks that were reported as impossible to attempt prior to surgery. While all patients continued to report feeling "moderately depressed" for several weeks, several also indicated that the sensations of "painful emptiness" and "void" remitted almost immediately following onset of stimulation at the optimal contacts.

Two patients failed to show a sustained antidepressant response at the 6 month time point (Table 2). One of these subjects, patient 2, met criteria for clinical response in the first 4 months; however, the level of improvement fluctuated over time, and the maximal benefit could not be recaptured with either a change in the stimulation contact or adjustments in stimulation parameters after 4 months. Patient 4 had no appreciable clinical improvement with chronic stimulation despite trials with various combinations of contacts and stimulation parameters. Of note, the prominent sleep disturbances in these two patients (difficulty falling asleep [patient 4] and hypersomnia [patient 2]) were not affected by DBS, unlike the middle insomnia improvements seen in the other four subjects.

Control Condition: Blinded DBS Discontinuation

We next considered that the long-term benefit in the responders could be related to placebo or nonspecific factors. To examine this possibility, after a period of continuous stimulation for 6 months, the effects of cessation and reintroduction of stimulation was examined in patient 1, who had shown the earliest, most robust, and best-sustained clinical response (Table 2). Following blinded discontinuation of bilateral stimulation (stimulators set at 0.0 V), antidepressant effects were maintained for 2 weeks (HDRS = 9; PANAS positive score = 48 of out of a possible 50, PANAS negative score = 10 out of a possible 50 versus 6 month score positive = 50, negative = 10). In weeks 3 and 4 without stimulation, the improvements in mood were also sustained (HDRS = 10). In the context of this sustained euthymia, however, there was a progressive change in behavior characterized by loss of energy and initiative, impaired concentration, and reduced activities, reflected by a drop in the PANAS positive score to 37, without appreciable change in the negative score (negative = 13). At this point, and under continued blinded conditions, the stimulator was turned back on to the previous best settings (3.5 V, PW 60, 130 Hz). This resulted in normalization of symptoms within approximately 48 hr and return to prediscontinuation activities within 1 week (HDRS = 6, PANAS positive = 50, negative = 10 after 1 week of restarting chronic DBS). This remission was sustained at 6 weeks of resu-

Table 3. Psychiatric Ratings: Patient Subgroups

	All Patients (n = 6) Mean Scores (SD)				Responders (n = 4) Mean Scores (Range)				Nonresponders (n = 2) Mean Scores (Range)			
	Preop	1 mo	3 mo	6 mo	Preop	1 mo	3 mo	6 mo	Preop	1 mo	3 mo	6 mo
HDRS 17	25.8 (2.8)	15.8 (4.7)	12.8 (7.8)	11.5 (6.8)	27.3 (2.1)	15.3 (5.4)	9.3 (5.9)	7.8 (3.1)	23 (22–24)	17 (14–20)	20 (15–25)	19 (15–23)
HDRS 24	34.6 (1.9)	25.8 (9.1)	21.2 (8.7)	18.8 (10.6)	35.7 (1.5)	25 (12.5)	15.3 (3.5)	11.3 (2.9)	33 (32–34)	27 (24–30)	30 (27–33)	30 (27–33)
MADRS	33.3 (4.5)	23 (8.5)	17.4 (10.1)	18.5 (10.4)	33.8 (6.7)	20.7 (11.1)	11 (3.6)	9.7 (3.8)	33 (31–34)	27 (25–28)	27 (21–33)	29 (25–33)
CGI	6.2 (0.4)	5.2 (0.8)	4.2 (0.6)	4.0 (1.7)	6.3 (0.5)	4.7 (0.6)	3.3 (1.0)	3.0 (0.8)	6 (6)	6 (6)	6 (6)	6 (6)

Note: f/u MADRAS and HDRS 24 scores not available for patient 6.

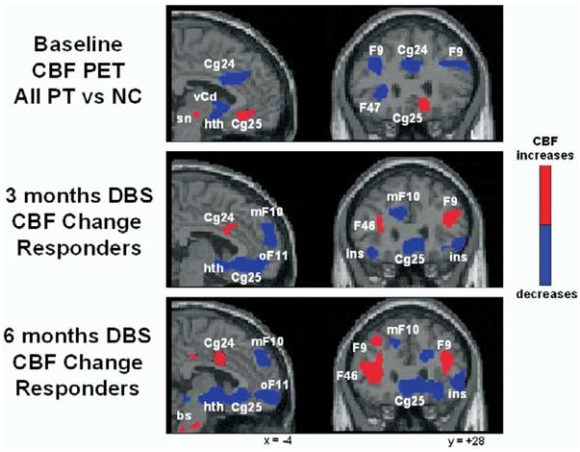


Figure 2. PET Scans

Regional cerebral blood flow changes (CBF PET) in TRD patients at baseline (row 1) and after 3 months (row 2) and 6 months (row 3) of successful treatment with continuous DBS. Sagittal (left) and coronal (right) views. Baseline CBF abnormalities are seen relative to age- and gender-matched healthy control subjects (NC): increases in subgenual cingulate (Cg25) and decrease in dorsolateral prefrontal (F9), ventrolateral prefrontal (F47) and anterior cingulate (Cg24) cortices (row 1, patients 1–5). Three months of DBS relative to baseline (row 2, patients 1, 3, and 5): decreases in Cg25, hypothalamus (Hth), anterior insula (ins), medial frontal (mF10) and orbital frontal (oF11); increases in prefrontal (F9/46) and dorsal cingulate (cg24). This same pattern is maintained at 6 months, although additional increases are seen in the brainstem (bs) (row 3). Slice location is in millimeters relative to anterior commissure. Numbers are Brodmann designations. L, left. Significant CBF increases in red; decreases in blue ($p < 0.001$).

med stimulation at comparable levels to the prediscontinuation baseline (HDRS = 4). Taken together, these findings suggest that stimulation of Cg25WM produces long-term improvements in mood that are sustained beyond the period of active stimulation. The cognitive aspects of depression (i.e., poor concentration, apathy) also show sustained improvements, but the observed changes appear to have a different biology and kinetics, decaying closer to the cessation of stimulation.

PET Measures of Regional Blood Flow: Baseline and Treatment Effects

Positron emission tomography was used to characterize the activity in brain networks involved in TRD and to provide a quantitative measure of brain changes associated with stimulation. Baseline, resting-state, cerebral blood flow (CBF) PET scans were performed in the first five study subjects and compared to five age- and sex-matched, nondepressed healthy volunteers. Depressed patients showed a unique pattern of elevated subgenual cingulate (Cg25) blood flow at pretreatment baseline, not previously reported in studies of nontreatment resistant patients. In addition, and consistent with past studies of depressed patients (reviewed in Mayberg, 2003), CBF decreases in prefrontal (BA9/46), premotor (BA6), dorsal anterior cingulate (BA24), and anterior insula were also identified (Figure 2 row 1; Table 4, left). A similar pattern of hyperactive Cg25 and hypoactive prefrontal cortex was seen in both the DBS responders and nonresponders (data not shown). Responder versus nonresponder differences at baseline were seen primarily in the magnitude of the prefrontal decreases (responders > nonresponders). Responders also showed an area of hyperactivity in the medial frontal cortex (BA10) not seen in the nonresponders; however, the small sample size precluded further analysis.

The time course of CBF changes with chronic stimulation of Cg25WM was also examined to establish the relationship of regional brain functioning to behavioral effects. Serial scans were performed after 3 and 6 months in four of the first five patients (1, 2, 3, and 5). Group analyses showed local CBF decreases in Cg25 and adjacent orbital frontal cortex (BA11) after 3 months of stimulation. The long-term responders (patients 1, 3, and 5) showed additional CBF changes at both 3 and 6 months: decreases in hypothalamus, anterior insula, and medial frontal cortex (BA10) as well as increases in dorsolateral prefrontal (BA9/46), dorsal anterior (BA24) and posterior cingulate (BA31), premotor (BA6), and parietal (BA40) regions (Table 4; Figure 2, rows 2 and 3). Neither the medial frontal (BA10) decreases nor the dorsal prefrontal (BA9/46), anterior cingulate (BA24), or parietal (BA40) increases were seen in the nonresponder (patient 2) at either 3 or 6 months (data not shown).

The stimulation-induced CBF increases in prefrontal cortex (BA9/46) normalized pretreatment abnormalities. Similarly, the Cg25 decreases not only normalized pretreatment dysfunction, but activity in this region with DBS was actually suppressed below that of the controls at both time points, a change also observed in our previous studies of PET scan changes with response to antidepressant medications (Mayberg et al., 2000). Unlike medication, brainstem changes were not seen early with DBS, although changes in the pons were demonstrated at the 6 month time point. Overall, regional changes seen after 3 months were maintained at 6 months in all three responders (Table 4, middle and right sections).

Neuropsychological Testing

A comprehensive neuropsychological battery was performed prior to surgery and after 3 and 6 months of stimulation, coincident with the day of the PET scan sessions. The test battery was designed to differentiate dorsolateral, superior medial, and ventrolateral/orbital frontal behaviors, potentially affected by chronic Cg25WM DBS. Serial testing further allowed differentiation between subacute surgical effects, chronic stimulation effects, and correlations with mood change.

At baseline, all patients were functioning intellectually in the average range, consistent with estimates of premorbid IQ. Inspection of results over time indicates that surgery itself did not have a negative impact on general cognition (i.e., IQ, language, basic visual-spatial function). Moreover, many specific areas that were below average or impaired at baseline were significantly improved (or trending due to low power) following 6 months of DBS [responders: visuo-motor function, particularly with the nondominant hand, $t(2) = 5.8$,

Table 4. PET Blood Flow Changes

Baseline[a] Patients (n = 5) vs Controls (n = 5)[c]						3 Months DBS vs Baseline[b,c] Responders (n = 3)						6 Months DBS vs Baseline[b,c] Responders (n = 3)					
Region	BA		Δ	z Score	x y z Coordinates	Region	BA		Δ	z Score	x y z Coordinates	Region	BA		Δ	z Score	x y z Coordinates
sgCg	25	L	↑	5.16	−10 28 −12	sgCg	25		↓	−4.75	−2 8 −10	sgCg	25		↓	−3.88	10 20 −4
Hth			↓	−3.67	0 4 −12							Hth			↓	−4.63	−2 2 −4
						OrbF	32/10		↓	−5.1	0 34 −8	OrbF	32/11	L	↓	−4.97	−10 30 −10
							32/10	R	↓	−4.84	6 46 2						
							10	R	↓	−3.98	22 60 −10		11/10	L	↓	−4.41	−20 36 −10
aIns		L	↑	6.05	−38 22 14	Ins		R	↓	−3.76	60 10 −6	Ins		R	↓	−4.07	58 16 −6
								L	↓	−5.74	−50 16 −16			L	↓	−5.67	−50 20 −6
mFr	10	R	↑	6.22	26 42 16	mFr	10	L	↓	−4.19	−22 62 22	mFr	10	L	↓	−3.53	−14 56 38
							10	R	↓	−5.95	34 56 26		10	L	↓	−4.27	−28 56 32
							24	R	↓	−5.86	12 22 26		24	R	↓	−4.19	12 20 20
													24	L	↓	−3.79	−16 24 22
						mFr	9	R	↓	−5.57	12 46 34	mFr	9	R	↓	−3.95	6 52 30
								L	↓	−4.87	−5 56 34		9	R	↓	−3.34	2 46 40
						Fr pole	9	L	↓	−5.96	−28 54 43						
								R	↓	−5.19	18 52 34						
pCg	31	L	↑	4.43	−8 −54 26	pCg	31	L	↑	4.65	−10 −72 20						
		R	↑	3.88	10 −56 26		31	R	↑	3.51	24 −68 14						
DLPF	9	L	↓	−7.07	−38 28 30	DLPF	9	L	↑	4.05	−26 16 30	DLPF	9	R	↑	4.63	38 12 34
		R	↓	−4.92	40 20 30								46/9	L	↑	3.79	−34 18 24
VLPF	47	L	↓	−4.66	−52 42 2		46	L	↑	4.73	−38 32 18		46	R	↑	4.13	32 24 10
		R	↓	−4.98	32 26 0		46	R	↑	4.91	34 22 16		46	L	↑	3.84	−38 26 10
dCg	24b	L	↓	−5.39	−2 18 28	dCg	24b	L	↑	4.06	−2 10 28	dCg	24b	L	↑	3.58	−4 4 34
						PM	6	L	↑	4.21	−52 −4 32	PM	6	L	↑	3.34	−48 2 22
							6	R	↑	5.71	50 0 30		6	R	↑	5.92	50 4 28
												Par	39	L	↑	4.13	−36 −56 18
													39	R	↑	4.01	44 −56 10
Caudate		R	↓	−4.69	14 2 12							PAG		R	↓	−4.66	6 −36 −10
aThal		R	↓	−4.58	12 −6 14												
n. acc		L	↓	−4.68	−4 10 −4												

[a] Patients 1–5 versus five age- and gender-matched healthy controls.
[b] Patients 1, 3, and 5.
[c] SPM t maps. Significance threshold: $p < 0.001$ uncorrected ($t > 3.27$); cluster size >50 voxels.

p = 0.014; dorsolateral frontal function (verbal fluency), t(2) = 10.0, p = 0.005; ventral prefrontal function (fewer errors on object alternation task), t(2) = 1.7, p = 0.12; and orbital frontal function (fewer risky choices on the gambling task), t(2) = 6.3, p = 0.012]. Importantly, there were no acquired impairments in orbital frontal functioning to indicate local DBS adverse effects (Kartsounis et al., 1991; Dalgleish et al., 2004). Nonresponders had normal performance on all tests at baseline, with the exception of slowed psychomotor speed (consistent with effects of depression). Repeat testing was only available for one of these subjects.

Adverse Events

Patients 2 and 4 developed local infections related to the connector cable at the chest (patient 2) or scalp (patient 4). Both were treated with intravenous antibiotics. Because of persistent infection in the absence of clinical benefit, the devices were explanted at approximately 6 months with resolution of their infections. No worsening of depressive symptoms was observed in either subject following explantation. Patient 5 developed skin erosion over the hardware and also received antibiotics. While no definitive etiology of these infections was identified, the protocol was modified after patient 5 to implant both the electrodes and pulse generator in a single surgical session, thus eliminating the several-day period where the electrodes remained externalized.

Discussion

In this study, we have demonstrated that high-frequency DBS of the Cg25WM can produce striking behavioral changes in patients with TRD. Acute behavioral effects were time locked to stimulation intraoperatively and during short-term testing sessions. Furthermore, sustained clinical improvements decreased with blinded discontinuation of chronic DBS and were recaptured with reinstitution of stimulation, providing evidence as to the specificity of DBS-mediated changes. PET scan data further indicate that Cg25WM DBS has profound effects on the cerebral networks involved in depression and suggest that reversal of baseline abnormalities correlates with antidepressant benefits. While the number of subjects is small, four of six patients (66%) achieved sustained clinical response or remission at the end of 6 months without changes in concurrent medications. This response rate is striking given the extreme treatment refractoriness of this patient population and the well-documented low and poorly sustained placebo response rates of such pa-

tients (Fava, 2003; Keller, et al., 1992; Kennedy and Lam, 2003; Thase and Rush, 1997; Sackeim et al., 2001).

The identified CBF change pattern associated with DBS-induced antidepressant effects has important similarities to other therapies for depression. The pattern of reduced activity in the Cg25 and hypothalamus together with prefrontal (BA9) and brainstem increases during DBS are identical to changes reported with antidepressant response to medication (Mayberg et al., 2000). On the other hand, the medial frontal (BA10)/orbital frontal (BA11) decreases and dorsal cingulate (BA24) increases are similar to the change pattern observed with response to cognitive behavioral therapy (Goldapple et al., 2004). The observation that the changes with DBS seen here features both the antidepressant medication and the CBT- induced changes suggests that Cg25WM DBS acts at a critical node of a distributed mood-regulatory network involved in major depression (Mayberg, 2003; Seminowicz et al., 2004).

The baseline pattern of subgenual cingulate hyperactivity in combination with frontal hypoactivity described here in this TRD patient group is a finding that is in contrast to the hypoactivity reported in a more rostral region of subgenual medial prefrontal cortex in familial bipolar and unipolar depressed patients (Drevets et al., 1997). This distinction suggests important differences across subtypes of depression that are potentially relevant to the pathophysiology of major depressive disorders and perhaps their treatment.

Although the mechanisms of action of DBS are incompletely understood, it is clear that stimulation produces both local and remote regional effects (Davis et al. 1997). The behavioral and neuroimaging changes in this study are consistent with suppression of the abnormally elevated baseline Cg25 activity (Figure 2). Possible mechanisms include DBS-induced activation of inhibitory GABA-ergic afferents and/or high frequency stimulation induced synaptic or metabolic failure (Gabbott et al., 1997; Lozano et al., 2002; McIntyre et al., 2004). The remote cortical and brainstem increases could occur as an indirect consequence of *trans*-synaptic effects in response to decreased activity in Cg25, or direct anterograde or retrograde activation of WM projections coursing through the stimulation field (Barbas et al., 2003; Carmichael and Price, 1996; Haber, 2003; Freedman et al., 2000; Ongur et al., 1998; Jurgens and Muller-Preuss, 1977; Vogt and Pandya, 1987).

In context of these putative mechanisms, the sudden mood change seen with stimulation in the operating room is consistent with an acute deactivation of a hyperactive Cg25, the region immediately adjacent to the area of stimulation. This region has been previously shown to mediate negative mood (increases with sadness, decreases with alleviation of depressive dysphoria) (Mayberg et al., 1999). Early effects of DBS on sleep, energy and motivation also suggest changes in activity of the hypothalamus and brainstem, regions monosynaptically connected to Cg25 (Barbas et al. 2003; Freedman et al., 2000; Jurgens and Muller-Preuss, 1977; Ongur et al., 1998). The more global changes in functioning, i.e., resolution of the depressive syndrome, as well as the prolonged washout time and associated persistent benefit we have observed beyond the cessation of stimulation, may reflect long-term changes in neural network properties as a consequence of prolonged stimulation (Carmichael and Price, 1996; Vaidya and Duman, 2001; Haber, 2003; Vogt and Pandya, 1987). This phenomenon of ongoing benefit after DBS discontinuation has been observed in other disorders targeting different brain circuits, including epilepsy, essential tremor, dystonia, and Parkinson's disease (Hodaie et al. 2002, Kumar et al. 2003; Deep-Brain Stimulation for Parkinson's Disease Study Group, 2001; Vidailhet et al., 2005).

Despite these encouraging results, there are limitations to this first study of Cg25WM DBS for TRD. Sample size was small, follow-up was limited, and no sham surgery or systematic placebo control arm was used. There was also inadequate power to identify demographic, clinical, subtype, neuropsychological, or imaging markers that might predict response. It may be relevant that all four responders had their first major depressive episode (MDE) before age 35 and had predominant melancholic features whereas the two nonresponders had first episodes in their 40s and had more atypical symptoms (Table 1). Differences in electrode targeting and placement and stimulation parameters may have also contributed to the observed response variance. The relative contribution of these various factors will require testing of additional subjects. In the future, the possible interactions between DBS and antidepressant medications need to be examined.

In conclusion, Cg25WM DBS appears to be efficacious for TRD. The treatment is titratable; there are objective imaging correlates of response effects; and the procedure is well tolerated. This approach may represent an effective, novel intervention for severely disabled patients with treatment-resistant depression.

Experimental Procedures

Patients

This pilot study included six patients with treatment-resistant major depression, referred by mood disorder specialists (Table 1). The clinical diagnosis of major depressive disorder, major depressive episode (MDD-MDE) was independently confirmed by two psychiatrists and a research coordinator using the Structured Clinical Interview for DSM-IV (First et al., 2001). Patients were selected for surgery because they were resistant to all available therapeutic options. All had failed to respond to a minimum of four different classes of antidepressant medications, prescribed at maximal tolerable doses. Failed treatments included SSRI, venlafaxine, bupropion, monoamine oxidase inhibitor, and tricyclic antidepressants, as well as augmentation strategies using lithium, atypical antipsychotics, and anticonvulsants. Five of the six patients had received electroconvulsive therapy, and all had attempted cognitive behavioral therapy without clinical improvement. Patients with cerebrovascular risk factors or a previous stroke; documented head trauma or neurodegenerative disorders; other Axis I psychiatric diagnoses including schizophrenia, bipolar disorder, panic disorder, obsessive compulsive disorder, or evidence of global cognitive impairment were excluded. Patients with psychotic symptoms, current history of substance abuse, daily use of alcohol in the previous 3 months, or active suicidal ideations were also excluded. Other exclusion criteria included age over 60, pregnancy (risk of radiation exposure from the PET scans), general contraindications for DBS surgery (cardiac pacemaker/defibrillator or other implanted devices), and inability or unwillingness to comply with long-term follow-up. Final selection was made by consensus of the investigator team. Additional subjects were screened but either did

not meet inclusion/exclusion criteria or were accepted but declined.

Informed Consent Process

The study protocol was reviewed and approved by three separate ethics committees at the University of Toronto: Toronto Western Hospital, Centre for Addiction and Mental Health, and Baycrest Centre for Geriatric Care. All potential subjects were referred by a psychiatrist that was not involved in the research protocol. They were then seen independently by two psychiatrists (V.V. and S.K.) for diagnostic assessment and if potentially eligible, introductory discussions took place. In all cases, the referring psychiatrist was involved in discussions of the potential benefits and the potential risks associated with DBS. During subsequent meetings with one of the principal investigators (H.S.M.), a family member was invited to attend and did so in all cases. Only after these discussions, were interested candidates referred to the neurosurgical principle investigator (A.L.), who explained the surgical procedure. Interested candidates were offered informed consent forms to read and further discuss as needed. The average time from initial assessment to signing consent was 6–8 weeks.

Surgery

The general surgical procedure for the implantation of DBS electrodes has been previously described (Abosch and Lozano, 2003). A stereotactic frame (Leksell G; Elekta, Inc., Atlanta, GA) was affixed to the patient's head on the morning of surgery, and preoperative MR images were obtained (Signa, 1.5 tesla; General Electric, Milwaukee, WI). The x, y, and z coordinates of the anterior (AC) and posterior commissures (PC) were determined using axial 3D T1 MR images. To target the subgenual cingulate white matter target, a midline T2 sagittal image was chosen, and the cingulate gyrus below the genu of the corpus callosum was identified (Figure 1, row 1) (Schaltenbrand and Wahren, 1977). A line was traced from the most anterior aspect (genu) of the corpus callosum to the anterior commissure and the midpoint was selected (Figure 1, row 2, left). The T2 coronal section correspondent to the plane of this midpoint was identified, and the coordinates of the transition between the gray and white matters of area 25 were calculated (Figure 1, row 2, right).

In the operating room under local anesthesia, a burr hole was drilled 2 cm from the midline in front of the coronal suture. The underlying dura mater was opened, and the exposed pial surface coagulated. Tisseal (Immuno, Vienna, Austria) was used to prevent cerebrospinal fluid egress and minimize brain shift. The Leksell arc was attached to the head frame and set to the target coordinates. Microrecordings were started 10 mm above the target using electrodes made from parylene-C-insulated tungsten wires and plated with gold and platinum. Tip lengths ranged from 15 to 40 μm and impedances ranged from 0.2 to 1.5 MΩ. Cell activity was amplified (DAM 80 WPI Instruments) with a gain of 1000 and initially filtered to 0.1–10 kHz. The signal was displayed on an oscilloscope and directed to a window discriminator (Winston Electronics) and an audio monitor (Grass AM 8, with noise clipping circuit). In the present study, microelectrode mapping was mainly used to confirm the anatomic location of the gray and white matters of area 25, characterized respectively by the recording of neuronal activity and cell sparse areas. The transition between these two regions was chosen as the final target for the implantation of the electrodes. Final electrode location was confirmed by postoperative MRI (example, Figure 1, row 3).

DBS quadripolar electrodes (Medtronic 3387; Medtronic, Inc., Minneapolis, MN) were implanted bilaterally. Each of the four electrode contacts was tested for adverse effects and clinical benefits. These contacts were numbered from 0 to 3 (right hemisphere) and 4 to 7 (left hemisphere), 0 and 4 being the most ventral and 3 and 7 the most dorsal contacts. The electrodes remained externalized for 5–7 days for clinical testing. They were then connected to a pulse generator (Kinetra, Medtronic, Minneapolis, MN) that was implanted in the infraclavicular region under general anesthesia. Prophylactic antibiotics were used for 24 hr after each of the surgical procedures.

Clinical Evaluation and Follow-Up

Clinical efficacy was evaluated using standardized ratings by the study psychiatrist blinded to the current stimulus parameters and/or changes. Standardized ratings included the Hamilton Depression Rating Scale (HDRS-17 and 24 item versions) (Frank, et al., 1991), the Montgomery Asberg Depression Scale (MADRS) (Montgomery and Asberg, 1979), the Clinical Global Impressions Scale (CGI) (National Institute of Mental Health, 1970), and the Positive and Negative Affective Scale (PANAS) (Watson and Clark 1988) (Table 3). Ratings were performed weekly for the first 3 months and biweekly until the study endpoint at 6 months, following baseline assessments at enrollment and 1 week prior to surgery. Medications were unchanged throughout the 6 month follow-up period.

PET Scanning Acquisition and Data Analysis

Regional cerebral blood flow PET scans (rCBF) were acquired preoperatively and after 3 and 6 months of chronic DBS (Fox et al., 1984). Five CBF scans were acquired in each subject at each time point. A comparative scan-set (one time point only) was also acquired under identical scanning conditions in a group of five age- and sex-matched healthy volunteers. All scans were acquired with subjects resting, with eyes closed and no explicit cognitive or motor instructions. Scans were acquired on a GEMS/Scanditronix 2048b camera (15 parallel slices; 6.5 mm center-to-center interslice distance) using measured attenuation correction (68Ge/68Ga transmission scans). rCBF was measured using the bolus [^{15}O]-water technique (35 mCi ^{15}O-water dose/scan; scan duration 60 s) (Mayberg et al., 1999). Scans were spaced a minimum of 11 min apart to accommodate radioactive decay to background levels. Mood state (sadness and anxiety) was assessed at the end of each scan using a seven-point analog scale and the PANAS to verify behavioral stability over the course of the five scans (Watson and Clark, 1988).

Statistical analyses were performed using SPM99 (Wellcome Department of Cognitive Neurology, London, UK) implemented in Matlab (version 5.3, Mathworks Inc., Sherborn, MA). PET scans were first realigned and then spatially normalized into standard three-dimensional space relative to the anterior commissure using the MNI ICBM 152 stereotactic template within SPM99. The images were then corrected for differences in the whole-brain global mean, and smoothed using a Gaussian kernel to a final in-plane resolution of 10 mm at full-width at half-maximum. Baseline abnormalities (patients versus controls) and changes with chronic DBS (3 months versus baseline, 6 months versus baseline) were assessed using general linear models (peak voxel threshold $p = 0.001$; cluster threshold 50 voxels (voxel = 2 mm^3) (Worsley et al., 1996). Resulting t values were lastly converted to z scores for interpretation. Brain locations are reported as x, y, z coordinates in MNI space with approximate Brodmann areas (BA) identified by mathematical transformation of SPM99 coordinates into Talairach space (http://www.mrc-cbu.cam.ac.uk/Imaging/) (Table 3).

Neuropsychological Testing

A comprehensive battery of neuropsychological tests was administered at three time points to establish baseline intellectual and cognitive abilities prior to surgery/stimulation and to monitor for changes over time (3 months, 6 months). Tests were chosen to tap general cognitive and intellectual function, as well as four domains of frontal function (Bechara et al. 1994; Freedman et al., 1998; Lang et al., 1999; Spreen and Strauss, 1998). Parallel versions were used where possible to minimize effects of repetition, and scores are corrected for effects of age, gender, and education, where appropriate. The following tests were administered: Wechsler Adult Intelligence Scale-III; North American Adult Reading Test; Trail Making Tests A and B; Boston Naming Test; Benton Judgment of Line Orientation Test; Hopkins Verbal Learning Test; Brief Visual Spatial Memory Test, Revised; Finger Tapping Test; Grooved Pegboard Test; Controlled Oral Word Association Test; Wisconsin Cart Sorting Test; Stroop Color Word Test; Emotional Stroop Task; Object Alternation Test; Iowa Gambling Task; and the International Affective Picture System Ratings. A subset of tests is presented. Paired t tests were used to compare differences between the baseline and 6 months data to determine the probability that the actual mean

difference is consistent with zero. This comparison is aided by the reduction in variance achieved by taking the differences, and thus is a good choice for use with a small sample. A more detailed discussion of neuropsychological results will be published separately.

Supplemental Data

The authors' conflict of interest statement can be found online at http://www.neuron.org/cgi/content/full/45/5/651/DC1/.

Acknowledgments

This study was supported by a distinguished Investigator Award to H.S.M. from the National Alliance for Research in Schizophrenia and Depression (NARSAD). We thank Kari Fulton for patient screening and coordination; Doug Hussey and Alvina Ng for acquisition of the PET studies; Jonathan Dostrovsky and William Hutchison for intraoperative electrophysiological studies; Jean Saint-Cyr for assistance with MR imaging; and Stanley Mayberg and Paul Holzheimer for critical review of the manuscript.

Received: January 4, 2005
Revised: February 4, 2005
Accepted: February 11, 2005
Published: March 2, 2005

References

Abosch, A., and Lozano, A.M. (2003). Stereotactic neurosurgery for movement disorders. Can. J. Neurol. Sci. *4*, S72–S82.

American Psychiatric Association, A.M. (2000). Practice guideline for the treatment of patients with major depressive disorder (revision). Am. J. Psychiatry *157*, 1–45.

Barbas, H., Saha, S., Rempel-Clower, N., and Ghashghaei, T. (2003). Serial pathways from primate prefrontal cortex to autonomic areas may influence emotional expression. BMC Neurosci. *4*, 25.

Bechara, A., Damasio, A.R., Damasio, H., and Anderson, S.W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. Cognition *50*, 7–15.

Benabid, A.L. (2003). Deep brain stimulation for Parkinson's disease. Curr. Opin. Neurobiol. *13*, 696–706.

Brody, A.L., Barsom, M.W., Bota, R.G., and Saxena, S. (2001). Prefrontal-subcortical and limbic circuit mediation of major depressive disorder. Semin. Clin. Neuropsychiatry *6*, 102–112.

Carmichael, S.T., and Price, J.L. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. J. Comp. Neurol. *371*, 179–207.

Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., and Poulton, R. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science *301*, 386–389.

Dalgleish, T., Yiend, J., Bramham, J., Teasdale, J.D., Ogilvie, A.D., Malhi, G., and Howard, R. (2004). Neuropsychological processing associated with recovery from depression after stereotactic subcaudate tractotomy. Am. J. Psychiatry *161*, 1913–1916.

Davis, K.D., Taub, E., Houle, S., Lang, A.E., Dostrovsky, J.O., Tasker, R.R., and Lozano, A.M. (1997). Globus pallidus stimulation activates the cortical motor system during alleviation of parkinsonian symptoms. Nat. Med. *3*, 671–674.

Deep-Brain Stimulation for Parkinson's Disease Study Group (2001). Deep-brain stimulation of the subthalamic nucleus or the pars interna of the globus pallidus in Parkinson's disease. N. Engl. J. Med. *345*, 956–963.

Dougherty, D.D., Weiss, A.P., Cosgrove, R., Alpert, N.M., Cassem, E.H., Nierenberg, A.A., Price, B.H., Mayberg, H.S., Fischman, A.J., and Rauch, S.L. (2003). Cerebral metabolic correlates as potential predictors of response to cingulotomy for major depression. J. Neurosurg. *99*, 1010–1017.

Drevets, W.C. (1999). Prefrontal cortical-amygdalar metabolism in major depression. Ann. N Y Acad. Sci. *877*, 614–637.

Drevets, W.C., Price, J.L., Simpson, J.R., Jr., Todd, R.D., Reich, T., Vannier, M., and Raichle, M.E. (1997). Subgenual prefrontal cortex abnormalities in mood disorders. Nature *386*, 824–827.

Fava, M. (2003). Diagnosis and definition of treatment-resistant depression. Biol. Psychiatry *53*, 649–659.

First, M.B., Spitzer, R.L., Gibbon, M., and Williams, J.B.W. (2001). Clinical Interview for DSM-IVTR (SCID-I): User's Guide and Interview–Research Version (New York: New York Psychiatric Institute Biometrics Research Department).

Fox, P.T., Mintun, M.A., Raichle, M.E., and Herscovitch, P. (1984). A noninvasive approach to quantitative functional brain mapping with H2 (15)O and positron emission tomography. J. Cereb. Blood Flow Metab. *4*, 329–333.

Frank, E., Prien, R.F., Jarrett, R.B., Keller, M.B., Kupfer, D.J., Lavori, P.W., Rush, A.J., and Weissman, M.M. (1991). Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Remission, recovery, relapse, and recurrence. Arch. Gen. Psychiatry *48*, 851–855.

Freedman, M., Black, S., Ebert, P., and Binns, M. (1998). Orbitofrontal function, object alternation and perseveration. Cereb. Cortex *8*, 18–27.

Freedman, L.J., Insel, T.R., and Smith, Y. (2000). Subcortical projections of area 25 (subgenual cortex) of the macaque monkey. J. Comp. Neurol. *421*, 172–188.

Gabbott, P.L., Jays, P.R., and Bacon, S.J. (1997). Calretinin neurons in human medial prefrontal cortex (areas 24a,b,c, 32′, and 25). J. Comp. Neurol. *381*, 389–410.

Goldapple, K., Segal, Z., Garson, C., Lau, M., Bieling, P., Kennedy, S., and Mayberg, H. (2004). Modulation of cortical-limbic pathways in major depression: treatment specific effects of cognitive behavioral therapy. Arch. Gen. Psychiatry *61*, 34–41.

Haber, S.N. (2003). The primate basal ganglia: parallel and integrative networks. J. Chem. Neuroanat. *26*, 317–330.

Hamilton, M.A. (1960). Rating scale for depression. J. Neurol. Neurosurg. Psychiatry *23*, 56–62.

Heim, C., Newport, D.J., Heit, S., Graham, Y.P., Wilcox, M., Bonsall, R., Miller, A.H., and Nemeroff, C.B. (2000). Pituitary-adrenal and autonomic responses to stress in women after sexual and physical abuse in childhood. JAMA *284*, 592–597.

Hodaie, M., Wennberg, R.A., Dostrovsky, J.O., and Lozano, A.M. (2002). Chronic anterior thalamus stimulation for intractable epilepsy. Epilepsia *43*, 603–608.

Hyman, S.E., and Nestler, E.J. (1996). Initiation and Adaptation: a paradigm for understanding psychotropic drug action. Am. J. Psychiatry *153*, 151–162.

Jurgens, U., and Muller-Preuss, P. (1977). Convergent projections of different limbic vocalization areas in the squirrel monkey. Exp. Brain Res. *29*, 75–83.

Kartsounis, L.D., Poynton, A., Bridges, P.K., and Bartlett, J.R. (1991). Neuropsychological correlates of stereotactic subcaudate tractotomy. A prospective study. Brain *114*, 2657–2673.

Keller, M.B., Lavori, P.W., Mueller, T.I., Endicott, J., Coryell, W., Hirschfeld, R.M., and Shea, T. (1992). Time to recovery, chronicity, and levels of psychopathology in major depression. A 5 year prospective followup of 431 subjects. Arch. Gen. Psychiatry *49*, 809–816.

Kendler, K.S., Thornton, L.M., and Gardner, C.O. (2001). Genetic risk, number of previous depressive episodes, and stressful life events in predicting onset of major depression. Am. J. Psychiatry *158*, 582–586.

Kennedy, S.H., and Lam, R.W. (2003). Enhancing outcomes in the management of treatment resistant depression: a focus on atypical antipsychotics. Bipolar Disord. *158*, 36–47.

Kumar, R., Lozano, A.M., Sime, E., and Lang, A.E. (2003). Long-term follow-up of thalamic deep brain stimulation for essential and parkinsonian tremor. Neurology *9*, 1601–1604.

Lancet UK ECT Review Group (2003). Efficacy and safety of

electroconvulsive therapy in depressive disorders: a systematic review and meta-analysis. Lancet *361*, 799–808.

Lang, A.E., and Lozano, A.M. (1998). Parkinson's disease. Second of two parts. N. Engl. J. Med. *339*, 1130–1143.

Lang, P.J., Bradley, M.M., and Cuthbert, B.N. (1999). International Affective Picture Rating System (IAPS): Instruction Manual and Affective Ratings. Technical Report A-4 (Gainseville, FL: The Centre for Research in Psychophysiology, University of Florida).

Lozano, A.M., Dostrovsky, J., Chen, R., and Ashby, P. (2002). Deep brain stimulation for Parkinson's disease: disrupting the disruption. Lancet Neurol. *1*, 225–231.

Maclean, P.D. (1990). The Triune Brain in Evolution: Role in Paleocerebral Function (New York, NY: Plenum).

Malizia, A. (1997). Frontal lobes and neurosurgery for psychiatric disorders. J. Psychopharmacol. *11*, 179–187.

Manji, J.K., Drevets, W.C., and Charney, D.S. (2001). The cellular neurobiology of depression. Nat. Med. *7*, 541–547.

Mayberg, H.S. (1994). Frontal lobe dysfunction in secondary depression. J. Neuropsychiatry Clin. Neurosci. *6*, 428–442.

Mayberg, H.S. (1997). Limbic-cortical dysregulation: a proposed model of depression. J. Neuropsychiatry Clin. Neurosci. *9*, 471–481.

Mayberg, H.S. (2003). Modulating dysfunctional limbic-cortical circuits in depression: towards development of brain-based algorithms for diagnosis and optimised treatment. Br. Med. Bull. *65*, 193–207.

Mayberg, H.S., Liotti, M., Brannan, S.K., McGinnis, S., Mahurin, R.K., Jerabek, P.A., Silva, J.A., Tekell, J.L., Martin, C.C., and Fox, P.T. (1999). Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. Am. J. Psychiatry *156*, 675–682.

Mayberg, H.S., Brannan, S.K., Mahurin, R.K., McGinnis, S., Silva, J.A., Tekell, J.L., Jerabek, P.A., Martin, C.C., and Fox, P.T. (2000). Regional metabolic Effects of fluoxetine in major depression: serial changes and relationship to clinical response. Biol. Psychiatry *48*, 830–843.

McIntyre, C.C., Savasta, M., Kerkerian-Le Goff, L., and Vitek, J.L. (2004). Uncovering the mechanism(s) of action of deep brain stimulation: activation, inhibition, or both. Clin. Neurophysiol. *115*, 1239–1248.

Mottaghy, F.M., Keller, C.E., Gangitano, M., Ly, J., Thall, M., Parker, J.A., and Pascual-Leone, A. (2002). Correlation of cerebral blood flow and treatment effects ofrepetitive transcranial magnetic stimulation in depressed patients. Psychiatry Res. *115*, 1–14.

Montgomery, S.A., and Asberg, M. (1979). A new depression scale designed to be sensitive to change. Br. J. Psychiatry *134*, 382–389.

National Institute of Mental Health, M. (1970). CGI: Clinical global impressions. In Manual for the ECDEU Assessment Battery. 2, W. Guy and R.R. Bonato, eds. (Chevy Chase, MD: National Institute of Mental Health).

Nestler, E.J., Barrot, M., DiLeone, R.J., Eisch, A.J., Gold, S.J., and Monteggia, L.M. (2002). Neurobiology of depression. Neuron *34*, 13–25.

Nemeroff, C.B. (2002). Recent advances in the neurobiology of depression. Psychopharmacol. Bull. *36* (Suppl.), 6–23.

Nobler, M.S., Oquendo, M.A., Kegeles, L.S., Malone, K.M., Campbell, C.C., Sackeim, H.A., and Mann, J.J. (2001). Decreased regional brain metabolism after ECT. Am. J. Psychiatry *158*, 305–308.

Ongur, D., An, X., and Price, J.L. (1998). Prefrontal cortical projections to the hypothalamus in Macaque Monkeys. J. Comp. Neurol. *401*, 480–505.

Sackeim, H.A., Rush, A.J., George, M.S., Marangell, L.B., Husain, M.M., Nahas, Z., Johnson, C.R., Seidman, S., Giller, C., Haines, S., et al. (2001). Vagus nerve stimulation (VNS) for treatment-resistant depression: efficacy, side effects, and predictors of outcome. Neuropsychopharmacology *25*, 713–728.

Schaltenbrand, G., and Wahren, W. (1977). Atlas for Stereotaxy of the Human Brain, Second Edition (Stuttgart, Germany: Thieme Verlag).

Seminowicz, D.A., Mayberg, H.S., McIntosh, A.R., Goldapple, K.K., Kennedy, S., Segal, Z., and Rafi-Tari, S. (2004). Limbic-Frontal Circuitry in Major Depression: A Path Modeling Metanalysis. Neuroimage *22*, 409–418.

Spreen, O., and Strauss, E.A. (1998). Compendium of Neuropsychological Tests, Second Edition (New York, NY: Oxford University Press).

Thase, M.E., and Rush, A.J. (1997). When at first you don't succeed: Seuqntial strategies for antidepressant nonrsponders. J. Clin. Psychiatry *58* (Suppl.), 23–29.

Vaidya, V.A., and Duman, R.S. (2001). Depresssion–emerging insights from neurobiology. Br. Med. Bull. *57*, 61–79.

Vidailhet, M., Vercueil, L., Houeto, J.L., Krystkowiak, P., Benabid, A.L., Cornu, P., Lagrange, C., Tezenas du Montcel, S., Dormont, D., Grand, S., et al. (2005). Bilateral deep-brain stimulation of the globus pallidus in primary generalized dystonia. N. Engl. J. Med. *352*, 59–67.

Vogt, B.A., and Pandya, D.N. (1987). Cingulate cortex of the rhesus monkey II: Cortical afferents. J. Comp. Neurol. *262*, 271–289.

Wang, P.S. (2003). National Co-morbidity survey replication. The epidemiology of major depressive disorder: results from the National co-morbidity Survey Replication (NCS-R). JAMA *289*, 3095–3105.

Watson, D., and Clark, L.A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. Personality Soc. Psychol. *6*, 1063–1070.

World Health Organization. (2001). Chapter 2: Burden of Mental and Behavioral Disorders. In The WHO Report 2001: Mental Health, New Understanding, New Hope. http://www.who.int/whr/2001/chapter2/en/Index3.html.

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., and Evans, A.C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp. *4*, 58–73.

Addiction (1993) 88, 1041–1054

THE SIXTH THOMAS JAMES OKEY MEMORIAL LECTURE

Vietnam veterans' rapid recovery from heroin addiction: a fluke or normal expectation?

LEE N. ROBINS

Department of Psychiatry, Washington University School of Medicine, St Louis, Missouri, USA

Abstract

Between 1972 and 1974, the outcomes of army enlisted men who had served in Vietnam during 1970–71 were evaluated and compared with that of a matched group. This paper reports the major findings of that study with respect to frequency of narcotic addiction in and after Vietnam, and the major risk factors for Vietnam addiction and later relapse. Extraordinary access to records facilitated drawing the sample, locating it, and verifying interview responses. The surprisingly low levels of readdiction and the rarity of addiction to narcotics alone as compared with poly-substance dependence are findings still not entirely incorporated into public and scientific views of heroin addiction. Some defenses against that incorporation are examined.

Introduction

When planning this paper, I was asked to look back at my study of Vietnam veterans (Robins, 1973, 1974b) to answer whether at the time I had thought it would be important, and to say how the results have fared over the last 20 years. I am grateful for the opportunity to indulge in this exercise of nostalgia.

The paper will begin by saying something about how the study happened to be done. Then it will summarize how the study turned out both technically and substantively. It will end by saying whether the findings seem to have been correct and generalizable, and if so, whether they have been integrated into the expert or public view of heroin addiction.

Why the Vietnam study was done

Vietnam was linked to the general concern in the United States about drug use from the time soldiers were sent there as 'advisers' in the early 1960s. First the concern was about marijuana. It was not until 1969 that heroin appeared in Vietnam. Rumors of its use by servicemen were confirmed dramatically in 1971 when two congressmen who travelled to Vietnam announced that many servicemen were addicted. Nixon then declared heroin addiction to be the nation's "No. 1 Public Health problem". He created the Special Action Office on Drug Abuse Prevention (SAODAP) in the White House, and tapped Dr Jerome Jaffe to head it. President Nixon persuaded Dr Jaffe to give up his academic life by arguing that wars had stimulated scientific discoveries in the past. Dr Jaffe accepted the argument, and agreed to take up the challenge of controlling heroin use in Vietnam, because if he failed, at least his efforts would yield a good scientific study. In June 1971, he went to Vietnam and established a urine screening program, which required that any man due to leave Vietnam be tested and have a 'clean' urine before

boarding the airplane. Urine initially positive by the FRAT test was to be re-tested with gas chromatography. If the test was again positive, the man was to be sent to detoxification for about a week, and his urine re-tested before he was allowed to leave the country. As soon as Dr Jaffe returned from Vietnam, he invited me to design a study to estimate the size of the problem both in Vietnam and after return.

My selection for the job was based on our having served together as members of the Narcotic Addiction and Drug Abuse Grant Review Committee of NIMH. I had been appointed to this committee based on a single published paper. My first major research effort was a follow-up of a child guidance clinic population and a matched normal control group (Robins, 1966). That study showed an impressive continuity between childhood antisocial behavior and adult outcomes, but I worried about the study's generalizability, because the subjects were patients, all white, and all born between 1907 and 1919. I decided to replicate it in a non-patient, later-born, non-white sample. In 1965–66, I interviewed 235 St Louis-born black men in their early thirties, selected from records of the St Louis elementary schools, which had been racially segregated in their childhoods. In both studies I looked at drug use and abuse as one outcome. While drug use was rare in the first study, the black schoolboys reached adolescence just after World War II, when drugs had entered the black ghettos but had not yet reached middle-class white America. By the time I followed them, illicit drug use had reached the white middle class and was a raging concern throughout the country. Half my sample said that they had used an illicit drug, and 10% said they had been addicted to narcotics. Although that was only 22 narcotics addicts, this was the first study of heroin addiction in a 'non-captive' adult population, i.e. neither in college nor in treatment. The paper reporting on their drug experience (Robins & Murphy, 1967) led to an invitation to serve on the Narcotic Addiction and Drug Abuse Grant Review Committee.

When Dr Jaffe invited me to do the Vietnam study, I was captivated. When I went to discuss it with him at the SAODAP office and found him working all hours, I began to realize the study might be important politically, as well as scientifically.

Why the study was a technical success

The study went well, helped by the government's concern about having exposed so many young men to what was viewed as a scourge.

It was possible to draw a sample representative of all enlisted Army men who left Vietnam in September 1971, the first month urine screening was in operation throughout the country. We selected two samples, one from a tape of 22 000 men made for us by the Department of Defense (DoD) from the computerized Active Duty Roster of November, 1971. That roster was purged every month of those discharged 4 months before. Thus, the November 1971 tape still contained all men in service in September. Men were selected if they were listed as leaving Vietnam that month. We were able to enrich the number of heroin users in the sample by sampling as well from records of those who tested positive at departure, records held in the Surgeon General's office. The military records of both groups were then checked to verify their departure dates. From Selective Service records, we were able to draw a sample of draft-eligible civilians who did not serve but matched the veterans with respect to age, region lived in, and education at the time the veterans entered service. The high level of cooperation from the subjects made possible a remarkably high interview completion rate, 96% when the veterans had been back 8–12 months, of whom 94% were then re-interviewed 3 years after their departure from Vietnam.

We also had access to all the record information we needed from the military and the Veterans Administration. These records, plus testing the urines we collected at the end of each interview, allowed us to verify what the men told us in interviews. These checks showed remarkable honesty: 97% of those whose military record showed narcotics use told us about their use while in service, and tests of urine samples collected at the end of the interview showed no higher rates positive for current use than did their self-reports given before they knew they would be asked for a urine sample.

We were able to complete the first interviews within a 6-month period, which both enabled us to provide information to the Government promptly, and meant that the men all had had approximately the same length of time at risk of relapse between return and interview. We reinterviewed the veterans and the comparison

sample when veterans had been back 3 years to learn whether the initial results would hold over time.

We began to feel confident about our sampling strategy when the DoD was not upset by our estimate that the number of men who returned in September 1971 was only 13 280, rather than the 22 000 on the tape given us. Our lower estimate was based on finding that many of the men we initially selected from the tape had to be dropped because their military records showed either that they never served in Vietnam or that they had left before September. While we were in the field, the DoD revised their estimate of men who returned in September based on lading documents from the airplanes that brought the men home, and we had hit within 3% of the revised number!

The study's assets

(A) A number of factors led to the high co-operation rate and the accuracy of our estimates. First, Dr Jaffe had involved multiple agencies—the National Institute of Mental Health, the Department of Defense, the Veterans Administration, and the Department of Labor—in its funding. This opened all the files we needed for sampling and follow-up.

(B) Urine testing in Vietnam had begun in July, 1971, 2 months before we designed the sample. This 2-month experience with testing had informed the DoD that the vast majority of heroin users were Army men below the officer level, and they advised us to restrict the study to this group.

(C) Being able to sample from the Surgeon General's list of men who had tested positive allowed us to over-sample heroin users so that we interviewed over 600 men who used narcotics in Vietnam. It also enabled us to estimate the proportion detected as drug-positive by the military at departure. Duplications between those we sampled from the Surgeon General's list and those we sampled from the Active Duty Roster gave us the needed data. Knowing the proportion positive then allowed us to weight the over-sample of positives back to their true representation in the population.

(D) All service records of those discharged happened to be stored in St Louis, making it convenient for us to review the complete military record to check whether the man had actually served in Vietnam and left in September, 1971. This turned out to be very important because the Army was in a 'stand-down' phase, during which manpower in Vietnam was being reduced by shortening the scheduled 1 year tour of duty there. Many of the names on our 'September' departure tape belonged to men who had actually left earlier. The military record also provided name and addresses of the next of kin, which helped us locate the man, as well as information about disciplinary actions and reported drug use, against which we could check honesty of the interview.

(E) Even though the men selected lived all over the US, and a few were stationed overseas, they were not difficult to find. Those who entered via the draft were drafted at 18 for 2 years. They typically arrived in Vietnam at age 19 after almost a year Stateside, served in Vietnam for 1 year, and were discharged soon after return, at 20 or 21. At this age, most could still be reached at their parents' home. The men who enlisted voluntarily had a 3-year obligation, and more of them signed up for another term of service. The military's Worldwide Locator made it possible to locate them at their current station.

(F) We had excellent support from the National Opinion Research Center, the group selected to carry out the interviewing. This was my first experience with employing an interviewing organization; on previous studies, I had hired and supervised interviewers myself. However, this study required rapid completion. An excellent study director and field director were willing to do extraordinary things, such as shipping interviews to Canada to preserve confidentiality through a double-link file, and collecting urines and shipping them to Canada for confidential testing. These extraordinary precautions to preserve confidentiality were vital to the enthusiastic cooperation of the interviewers, who were very concerned about possibly harming the veterans. (The men themselves seemed remarkably unconcerned.)

(G) A delay in funding, that looked ominous initially, turned out to be helpful in the end. Dr Jaffe had originally asked that we interview the men when they had been back 2–3 months. The funding delay meant that they had been back 8 months before we could enter the field. This delay gave them time to be discharged and return home, making them easier to find. It also made the findings more meaningful. If they

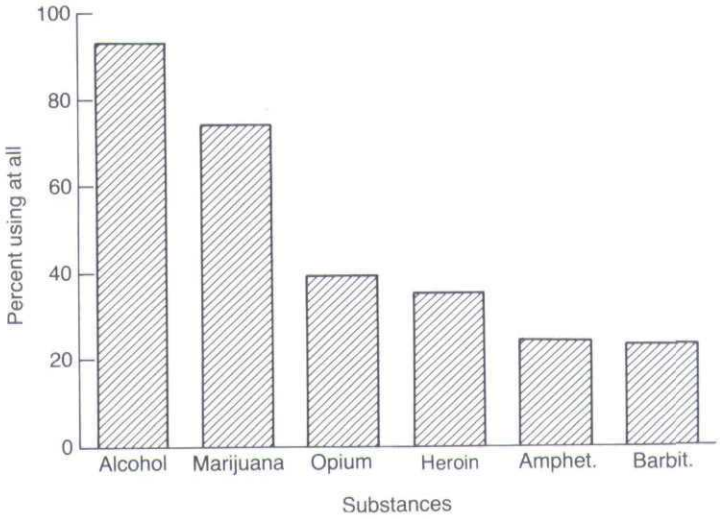


Figure 1. *Substance use in Vietnam.*

wanted to use heroin in the States, they had had ample time to locate a supply and begin to use it.

(H) Before the 3-year follow-up was due, the draft ended and selective service registrations were moved to a few Federal Records Centers. This made the location of civilian control subjects a great deal more efficient than if we had had to approach the Draft Board in each subject's home town to find a civilian who matched him. The DoD even provided us with members of Army Reserve Units, who spent their required yearly 6 weeks of active duty in going to the Federal Records Office to select civilian matches for our second follow-up. This overcame the problem that as civilians we could not legally have access to Selective Service records (although we accompanied the Reservists to be sure they selected cases properly.)

I believe these were the reasons that the study went well, along with the veterans' wish to be heard. But a technically successful study is exciting only if the results are interesting. Let me summarize what I see as the five major findings.

The major findings

(1) First, the DoD had greatly underestimated the size of the narcotics problem in Vietnam. Almost half (45%) of Army enlisted men there in 1970–71 tried narcotics; 34% tried heroin and 38% tried opium. While fewer used heroin or opium than alcohol or marijuana (Fig. 1), narcotics were used at a truly astounding rate. These figures reflect the fact that these drugs were cheap—even an addict need spend only about $6.00 per day—and pure—they need not be injected, a method which repelled many soldiers. Instead heroin could be mixed with tobacco and smoked. That opium was used so widely was a particular surprise; the press had reported only on heroin use.

Addiction was also more common than reported—20% claimed that while in Vietnam they had felt strung out or addicted to narcotics. To see whether that claim really meant physiological addiction, we asked for frequency and length of use and withdrawal symptoms. Those who claimed addiction had almost all used narcotics heavily for a considerable time and suffered the classic symptoms of withdrawal for at least several days. We concluded that their claims of addiction were correct.

Almost 11% of Army enlisted men's urines tested positive at departure. If the urine screen captured only those so addicted they could not stop their use when warned that they would be screened, as Dr Jaffe intended, this would mean that about 1400 recently or still addicted enlisted soldiers were arriving in the US per month (plus a small number of addicted officers and enlisted personnel in other branches of the military).

Serving in Vietnam was associated with greatly increased use of marijuana as well as narcotics; close to 80% of enlisted men used marijuana while in Vietnam. Other drugs were used too, but it was use of marijuana and narcotics that

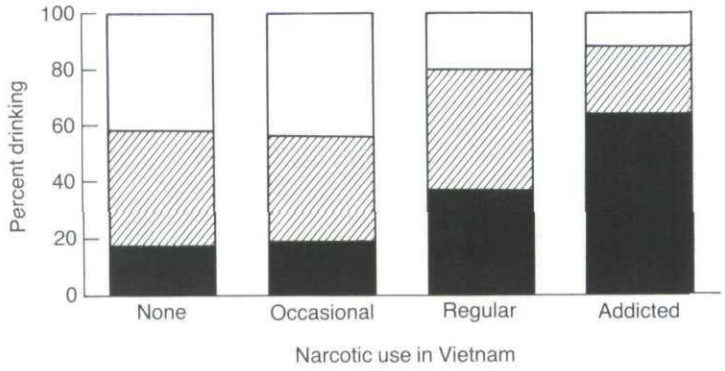


Figure 2. *In Vietnam, narcotic use and drinking were inversely related (Drinking frequency: ■ less than monthly; ▨ less than daily; □ daily).*

was distinctively greater in veterans than in the comparison sample.

(2) A second finding was that the orderly sequence of substance use described as a Guttman scale, in which no one progresses to an illicit drug without having used the legal drugs, alcohol and tobacco, and no one progresses to a 'hard' illicit drug without having used a 'softer' illicit drug (Kandel, 1975), had been turned topsy-turvy in Vietnam. Before and after Vietnam, heavy alcohol use was almost a prerequisite for narcotic use; in Vietnam, they were inversely correlated (Fig. 2). (This may have been explained by military rules against selling alcohol to soldiers under 21; for the average enlisted man, who arrived in Vietnam at 19, heroin was more available than alcohol.) This displacement from the expected sequence also occurred for amphetamines. Before service, amphetamines were often used without narcotics or barbiturates, but narcotic use virtually never occurred without amphetamine use. In Vietnam, amphetamines were essentially used *only* by users of barbiturates and narcotics. These sequence reversals show that a drug's position in the sequence has more to do with its availability than either with its intrinsic 'hardness', by which I think we generally mean its addiction liability, or with popular beliefs about its dangerousness.

(3) In my study of young black men, I had found a history of deviant behavior strongly associated with drug use (Robins & Murphy, 1967). One might have expected the much greater availability of drugs in Vietnam and stresses of serving in a war to break the link with deviance. But they did not. The same relationship between early behavior problems and drug use held in Vietnam. We counted occurrence and severity of five pre-service behaviors: fighting; truanting; drunkenness; arrest; and school expulsion. Each was scored as not present, present mildly, or present at a serious level. Both a larger number and greater seriousness increased the risk of using narcotics in Vietnam (Fig. 3).

(4) In the first year after return, only 5% of those who had been addicted in Vietnam were addicted in the US (Fig. 4; Robins, Davis & Nurco, 1974b). This finding was totally unlike the outcomes of young men treated in Lexington, the Federal Narcotics Hospital at the time. When those young men were followed 6 months later, two-thirds were found to already be re-addicted (Stephens & Cottrell, 1972). Even veterans still on narcotics at departure from Vietnam were doing very well 8–12 months after return. The curve showing the likelihood of any use, heavy use, and addiction for veterans was the mirror image of that for treated civilians. Nor was this good result transient. When we followed veterans at 3 years, only 12% of those addicted in Vietnam had been addicted at any time in the 3 years since return, and for those re-addicted, the addiction had usually been very brief.

It was not treatment that explained this remarkable rate of recovery. Only a third of the men addicted in Vietnam received even simple detoxification while in service, and only a tiny percentage of Vietnam enlisted men went into drug abuse treatment after return—less than 2% of those who used narcotics in Vietnam, 6% of those who were positive at departure, and 14% of those positive at departure who continued to

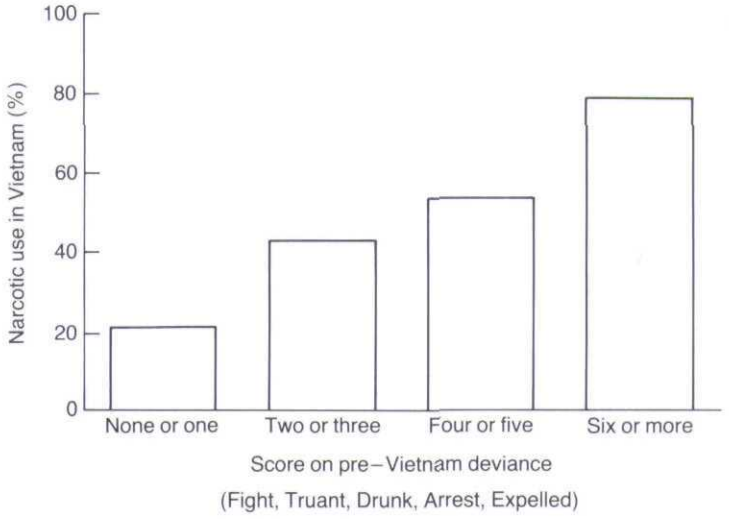


Figure 3. *Behavior before Vietnam and narcotic use in Vietnam.*

use after return (Robins, 1975). Yet, those who *did* enter treatment had relapse rates as high as the young civilian men in Lexington—two-thirds had relapsed by the time we interviewed them. Relapse often occurred the very day they left the hospital (Fig. 5).

Nor did recovery require abstention. Although nearly half the men addicted in Vietnam tried narcotics again after return, only 6% overall got re-addicted (Fig. 6). Some were spared by using only narcotics other than heroin; some by not injecting, some by using only occasionally. But even regular heroin users became re-addicted in only half the cases (Robins *et al.*, 1980).

This surprising rate of recovery even when re-exposed to narcotic drugs ran counter to the conventional wisdom that heroin is a drug which causes addicts to suffer intolerable craving that rapidly leads to re-addiction if re-exposed to the drug.

(5) In the US, many drug programs have been set up especially for heroin addicts, for whom methadone, experimental narcotic antagonists, or group therapy is provided. We found little evidence for a group of veterans whose main problem is heroin addiction (Robins *et al.*, 1978). Those who used heroin after return were also using a great variety of other drugs. More than 80% used amphetamines, more than 70% used barbiturates, and almost all used marijuana. Those addicted to heroin often had multiple dependencies. When heroin addicts were asked what their 'main' drug was, more than half named alcohol or marijuana rather than heroin.

We found little to justify the view of heroin as an especially dangerous drug. Heroin *was* associated with more adverse outcomes (crime, unemployment, illegal employment, divorce or separation, violence, transiency, credit problems, alcohol abuse, drug problems) than barbiturates and amphetamines, but *not if pre-service behavior problems and the number of other types of drugs used was held constant.* When they were, barbiturates and amphetamines were associated with as many problems. Indeed, the *variety* of drugs used was a better predictor of adverse consequences than *which* drugs were used.

These then were the major conclusions: narcotic use and narcotic addiction were extremely common in Vietnam (although not as common as use of alcohol and marijuana); availability was the main explanation; those with a history of deviant behaviour *before* Vietnam were particularly at risk; addiction was rare and brief after return, even when men continued to use narcotics; veterans re-addicted *and entering treatment* had as high a relapse rate as civilians; and heroin's adverse effects were no greater than those of amphetamines or barbiturates when juvenile behavior and concomitant use of other drugs was taken into account.

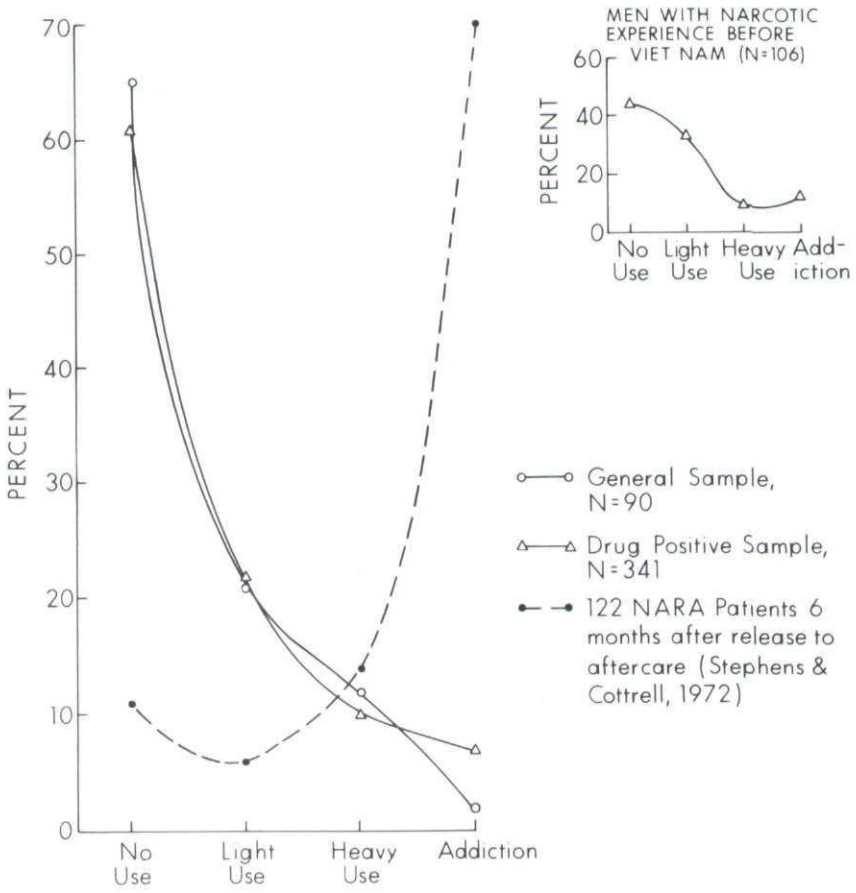


Figure 4. *Narcotic use in the 8–10 months post Vietnam by men first addicted in Vietnam.*

The response to the study by the DoD, the press, and the research community

Even though our study showed that the amount of heroin use in Vietnam had been underestimated, the DoD was pleased with the findings, because they showed that Vietnam veterans had not been consigned to a life of unrelenting dependence on drugs. (Our only problem was with one general who wanted to take credit for having cured the addicts with a little Valium during detoxification, and claimed that our study proved that methadone maintenance was unnecessary.)

The press and the research community were more skeptical. They resisted giving up the beliefs that heroin was a uniquely dangerous drug, to which a user became addicted very quickly, and addiction to which was virtually incurable. They maintained these beliefs by offering three interpretations of our findings:

(1) The results could be wrong. Perhaps the findings were tailored to exonerate the DoD; the men might have been addicted after return more often than we reported. (The suspicion that the results were politically manipulated may have been encouraged by the DoD's insistence on sponsoring the press conference which announced initial results.)

(2) Addiction in Vietnam was explained by the extraordinary setting—the misery of war that made addiction a 'normal' reaction; so the relatively benign outcome of addiction in Vietnam was irrelevant to addiction in the US.

(3) The drop in addiction on return was

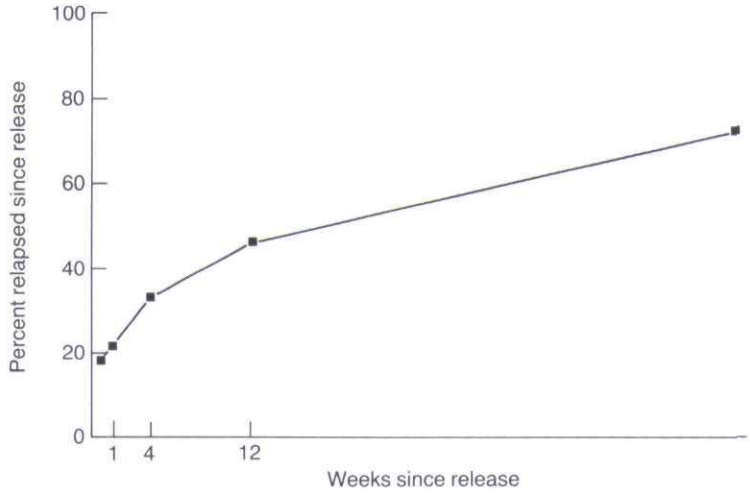


Figure 5. *Relapse to drug use following VA in-patient treatment.*

caused by a change in setting. In the US these men would not know where to get heroin, and the settings in which they lived and worked would not be associated in their minds with use or withdrawal symptoms, and therefore would not serve as stimuli to relapse. Thus again, the study findings would be irrelevant to addiction beginning in the US.

I will argue that each of these views fails to fit the facts.

(1) The idea of a whitewash by the DoD was extensively investigated by a New York Times (NYT) reporter who had recently returned from the newspaper's Saigon office. He was following up on a NYT news story in which a number of scientists had commented on the study, some with doubts about the results. During the 2 months' preparation of his article, he went over the study in great detail, often calling me to challenge what he took to be discrepancies between statements in my report. I spent a lot of time explaining how the results had been reached, and that percents *should* differ when the group being described shifts from the whole population of enlisted men to sub-populations, such as men who had already used drugs before arriving in Vietnam. When the article finally appeared, the Vietnam study was barely mentioned in it. To get only one line in the two-page story that had been intended as an exposé showed that he could not support suspicions of a DoD whitewash.

Of course, the idea that there was a high rate of addiction after return was even more definitively disproved by the failure of veterans to apply for treatment. One of the original motivations for the study was the VA's concern that returning addicts would overtax their services; but the anticipated large demand never occurred. A small increase in cases occurred only because of a change in DoD regulations a few months after the men's return, allowing the Army to delay discharge by 30 days, which drug-positive soldiers would spend at a VA drug treatment program. The DoD wanted men to become familiar with the VA system so that they would return to it if they relapsed. However, they arrived at the hospital already detoxified, and in no obvious need of treatment. The VA generally transferred them to the hospital nearest their homes and allowed them to go home on leave, while remaining on the hospital books as patients, as required by the regulation.

(2) The argument that addiction in Vietnam was a response to war stress, and therefore remitted on exit from the Vietnam war theatre, is still frequently cited as though it were self-evident, because it sounds so plausible. Yet accepting this argument is difficult in the face of the facts: Heroin was so readily available in Vietnam that more than 80% were offered it, and usually within the week following arrival. Those who became addicted had typically begun use early in

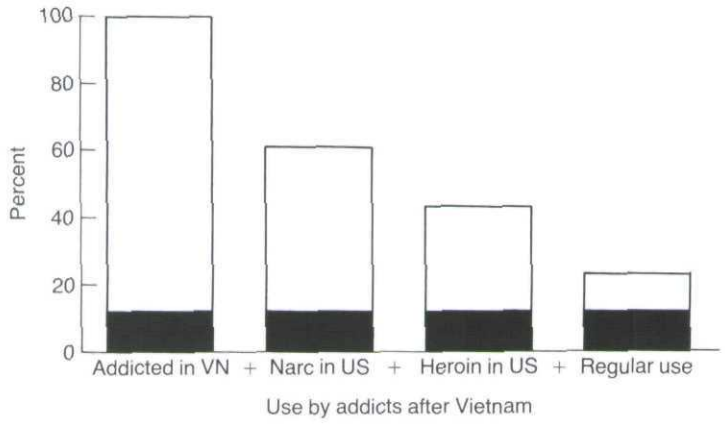


Figure 6. *Risk of re-addiction depending on degree of exposure (■ re-addicted; □ not re-addicted).*

their Vietnam tour, before they were exposed to combat. Further, the dose-response curve that is such a powerful causal argument did not apply: those who saw more active combat were not more likely to use than veterans who saw less, once one took into account their pre-service histories. (Those with pre-service antisocial behavior both used more drugs and saw more combat. Their greater exposure to combat was presumably because they had none of the skills that kept cooks, typists, and construction workers behind the lines.)

While combat had little effect, deviant behavior before service was a powerful predictor, as we have seen. So was pre-service drug experience (Fig. 7); the greater the variety of drugs used before entering service, the greater the likelihood that narcotics would be used in Vietnam. Like the St Louis-born young black men we studied earlier, the best predictors for use of narcotics in Vietnam were earlier antisocial behavior and experience with precursor drugs.

Finally, when we asked men why they used heroin, they did not tell us that they were overcome by fear or stress. Rather, they said it was enjoyable and made life in service bearable (Table 1).

(3) The argument that men addicted in Vietnam would not relapse because they would not be re-exposed in the US to stimuli associated with their drug use and withdrawal symptoms experienced abroad is based on Abraham Wikler's conditioned response theory of relapse (Wikler, 1973). This theory has been the basis for successful treatment programs, which gives it empirical support. However, it fails to account for the fact that among the one-half of the Vietnam addicts who *did* use heroin after their return, most did not become re-addicted, and those who did become addicted were addicted only briefly. Only one-fourth of those re-addicted by the time of their first interview at 8–10 months after return were addicted at any time in the next 2 years, even though, having been addicted in the US, they presumably had the opportunity to develop conditioned responses to the US stimuli associated with addiction and withdrawal (Fig. 8). Their claim not to be still addicted was validated by finding few of them to have positive urines at either the first or the second follow-up interview.

Where to the Vietnam study's findings stand now?

What then are my views of heroin addiction twenty years after the Vietnam study? I have still not found a serious flaw in the study, and I think the findings are consistent with those of other studies. For example,

(1) Charles Winick (1962), using the Federal Bureau of Narcotics records and the New York Narcotics Register, found that names tended to drop off at ages 35–40, or after 5 years. Because he carried out no personal interviews, he could not prove that his critics were wrong when they argued that the addicts had left town or died or somehow managed not to get caught again. However, he believed that they had 'matured out'—i.e. recovered spontaneously.

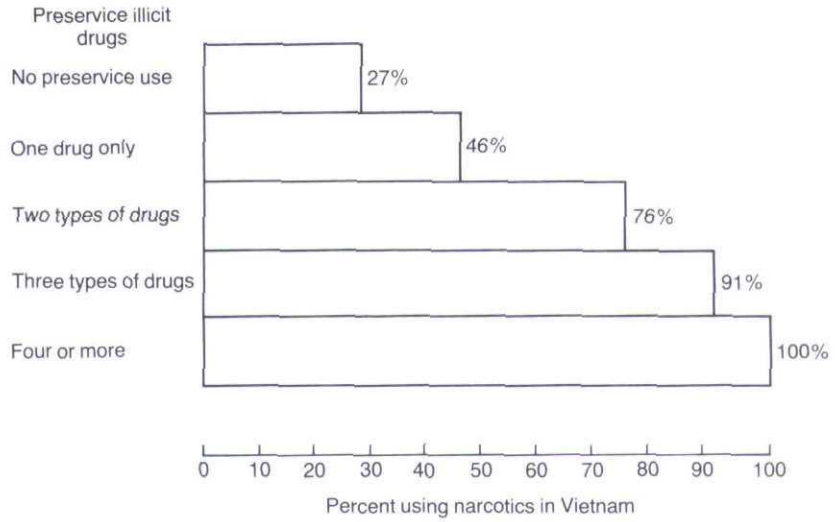


Figure 7. *Preservice drug experience and use of narcotics in Vietnam.*

(2) Waldorf & Biernacki (1979) did an ethnographic study of 50 untreated heroin addicts who had been free of addiction for at least 2 years, and reported that the median duration of their addiction had been only 5.7 years.

(3) Our own small study of young black men (Robins & Murphy, 1967) included 22 who reported having been addicted to heroin. Only four of these addicts had used any heroin at all in the year prior to interview at age 31–35. And only four had had any treatment. Heroin addiction for them, as for Vietnam veterans, was associated with use of a large number of other drugs. Like the Vietnam veterans, they experienced an accretion of drug types, not a 'stepping stone' pattern of moving from milder to stronger drugs. Indeed, heroin use preserved, rather than replacing, the use of other drugs.

(4) A study of young men 18–28 registered with Selective Service (O'Donnell *et al.*, 1976), conducted at the same time as our 3-year follow-up of veterans, showed that only 13% of those reporting heroin use had had any treatment, and only 27% of the untreated had used heroin within the last year, while 65% of the treated had. This suggested that treated cases are those who cannot stop on their own—and often do not stop when treated.

Table 1. *Why men used narcotics in Vietnam*

Reason volunteered	%
To feel high	40
Tolerate Army regulations	13
Relieve boredom	9
Relieve depression	9
Relieve fear	8
Pass time	5
Be one of the group	3

(5) A follow-up of Harlem youth (Brunswick, 1979) showed that only 25% of young people 18–23 who had used heroin twice or more had used any in the current year; the rate of recent use was particularly low among the untreated—16%.

(6) Norman Zinberg identified a group of 'chippers'—recreational users of heroin who never became addicted (Zinberg & Jacobson, 1976).

(7) The Epidemiologic Catchment Area study of all major psychiatric disorders in random samples of US adults in five geographic areas asked all affected the age of first symptom and the age of most recent symptom, and whether they had ever discussed their illness with a physician (Robins, Locke & Regier, 1991). Subtracting age of first problem from age at the last problem gives an estimate of duration. Substance abuse had the shortest mean duration of any of the psychiatric disorders—only 2.7 years. And only one in five of those dependent on or abusing an illicit drug had had any treatment. Most of these cases were not heroin addicts; marijuana was the most common drug of abuse. Still, this study

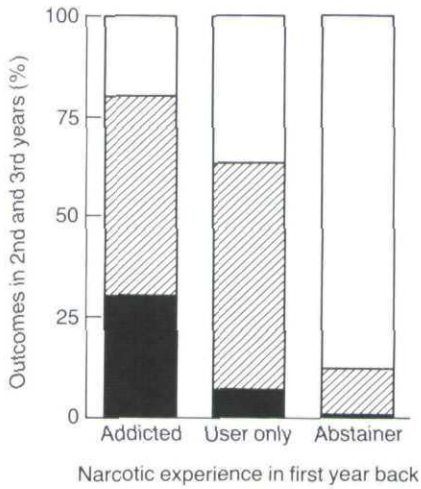


Figure 8. *Fate of year 1 addicts, users, and abstainers in the next 2 years (■ addicted year 2–3; ▨ user, non-addict; □ abstainer, year 2 and 3).*

shows that drug addiction tends to remit quickly, even in the absence of treatment.

So, I am unrepentant. I think the results were both correct and not unique. This is not to deny that the Vietnam situation was one quite unknown previously. Surely there has never been a time when so many young men were addicted to opiates in the absence of typical risk factors. Few of the Vietnam addicts would have become addicted if they had remained in the US. However, their history of brief addiction followed by spontaneous recovery, both in Vietnam and afterwards, was not out of line with the American experience; only with American beliefs. Nor was their ability to reuse without becoming re-addicted very unusual.

I think there are three important implications for policy in this study, which still have not been fully incorporated into our understanding of heroin addiction. First, addiction looks very different if you study it in a general population than if you study it in treated cases. The small number of Vietnam addicts who came to treatment relapsed as rapidly as do other young men in treatment. Their having become originally addicted as a response to stress or in an exotic situation was of no help to them. Those who did not come to treatment were able to get themselves off heroin without help. The implications are first, that heroin addicts are not doomed for life, and should be helped to remain in or re-enter conventional society, and second, that drug users who appear for treatment have special problems that will *not* be solved just by getting them off drugs.

Second, I think the evidence that heroin is not an exceptional drug has stood the test of time. Its association with long-term addiction and crime appears to be as much a function of its bad reputation, that relegates its use chiefly to people with little regard for social proprieties, as to its intrinsic effects on brain and behavior.

Third, at least in the US, the 'heroin addict' for whom current treatment programs are designed is a mythological creature. Most users of heroin use many substances, and may be as or more dependent on one of these other drugs as on heroin. They are almost all (in the US) also dependent on alcohol and tobacco. Yet treatment slots are still mainly available for heroin use and heroin users. In the Epidemiologic Catchment Area study conducted in the early 1980s, we found that persons with heroin problems in the last year had a reasonable chance (35%) of being seen in a drug treatment facility. Those with problems with other drugs only were very unlikely to receive treatment in a drug program (less than 2% did). Yet they were only a little less likely than those with heroin problems to receive some kind of psychiatric care (Fig. 9). The segregation of treatment by drug type seems misguided to me. It is no wonder that drug

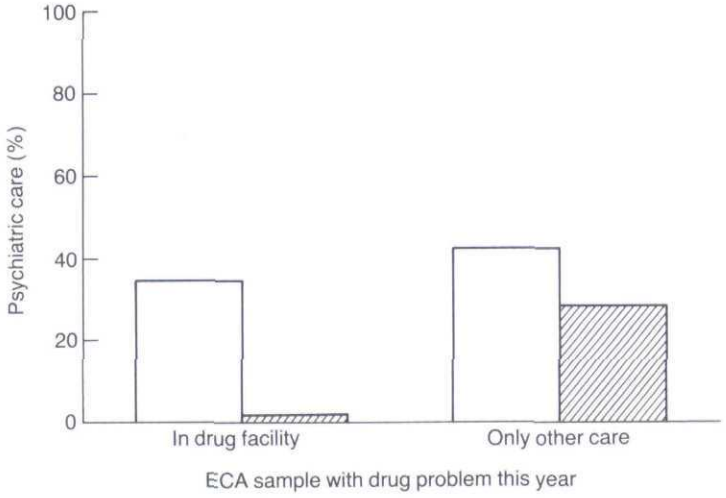


Figure 9. *Treatment this year; heroin (□) vs. other illicit drugs only (▨).*

treatment is not more successful when only a small part of the total problem is being addressed.

As I look back at the Vietnam heroin experience, I conclude that soldiers had no special readiness to use narcotics or to recover from addiction to them. Their remarkable rate of use was a response to market conditions—both the high availability of opiates and the lack of alternative recreational substances, to the absence of disapproving friends and relatives, and to the fact that serving in Vietnam was not seen as part of their real-life career. Their readiness to recover from addiction did not differ from that of other users. The reason that the press and scientists alike were surprised was because studies of the general population's drug dependence were and continue to be so rare. Their expectations were based on the rates of relapse found in patient populations, made up of addicts who tried but failed to get themselves off drugs.

It is not likely that the lack of general population studies will be remedied. Even large studies of general populations, such as the National Institute of Drug Abusers regular National Household Survey of Drug Use and yearly Monitoring the Future, which follows high school seniors into adulthood, do not study addiction because there are too few cases. The Epidemiologic Catchment Area study, committed to the study of disorder, not just use, reported on drug abuse and dependence as a whole because only 0.7% of the population met criteria for opiate abuse or dependence.

Beliefs about heroin based entirely on results in treated populations have created a self-fulfilling prophecy. Heroin is or was (perhaps now surpassed by 'crack') thought of, by law enforcement personnel and users alike, as the 'worst' drug, virtually instantly and permanently addictive and creating craving so extreme that it overcomes all normal ability to resist temptations to theft and robbery to acquire it. Users who share that view show by their use that they are ready to commit themselves to their concept of an addictive life style. The public's ranking of drugs with respect to 'hardness' probably has more to do with the drug's legal status, the government's commitment to discouraging its use, and its price than with any intrinsic addictive liability. This is best demonstrated by tobacco, a highly addictive drug ranked as 'soft'.

The fact that heroin addiction is not always interminable is still being 'discovered', even by scientists. Just last year the Neuroscience Institute at Harvard University published an article entitled "Addiction: What Science Knows" (Hyman, 1992). It reassured readers that *craving* does not condemn the user to indefinite use: "In the past, it was thought that physical withdrawal symptoms appearing with drug cessation were the key indicators of drug addiction . . . (Now, i)ndeed, despite the 'common wisdom', avoidance of physical withdrawal appears to be a

Table 2. *Variables that failed to predict re-addiction among veteran addicts who returned to narcotics*

Behavior problems before service
Parents' substance abuse, arrest, psychiatric care
Substance use in Vietnam (types, duration, injection)
IQ
Draftee or volunteer
Combat
Military discipline problems
Demographic characteristics

relatively minor factor in maintaining addiction even for . . . opiates . . ." But the article still supported the view that addiction is largely irreversible. Accepting *this* 'common wisdom' triggered straining after some biological basis for an irreversibility that may well not exist: "What causes the compulsive use? . . . (O)piates tap into (a) critical circuit in the ventral tegmental area . . . that interacts with the nucleus accumbens and . . . the cerebral cortex . . ., the 'brain reward pathway' . . . With repetitive drug use the brain reward system . . . undergo(es) adaptive changes involving . . . dopamine . . . When the drug is taken away, these neurons now function abnormally, producing dysphoria and intense drug craving . . . for many years." The author is too sophisticated to think that dopamine is the *whole* story: "Factors relevant to individual vulnerability include . . . heredity, . . . unknown developmental factors, . . . other psychiatric disorders, . . . pain, . . . stress. Sociocultural factors also . . . modify the risk of addiction, includ(ing) drug availability, and pressures for or against drug use in one's family, subculture or peer group."

Such a complicated explanation for a belief that addiction liability is relatively permanent is very difficult to test, particularly when the facts indicate that such permanence is rare rather than universal. The solution to achieving some simplification may lie in shifting our focus from addiction as a unitary phenomenon to the transitions between use, addiction, and recovery. In the Vietnam study, the social and behavioral variables that had been so useful in explaining who would *use* narcotics enough to become addicted to them in Vietnam and which ex-addicts would use them again after they returned to the US were of no use at all in explaining which relapsers would become re-addicted (Table 2). Vulnerability to re-addiction *if re-exposed* may indeed be biological. We don't know whether that biological vulnerability is present from birth or created by action of the addictive substance on the brain, as the Newsletter article suggested. What we *do* know is that *different* factors operate at each of the critical transitions: from the drug naive state to use; from use to addiction; from addiction to remission; from remission to relapse.

References

Bibliography of principal references from the Vietnam study

GOODWIN, D. W., DAVIS, D. H. & ROBINS, L. N. (1975) Drinking amid abundant illicit drugs: the Vietnam case, *Archives of General Psychiatry*, 32, pp. 230–233.

HELZER, J. E. & ROBINS, L. N. (1985) Specification of predictors of narcotic use vs addiction, in: ROBINS, L. N. (Ed.) *Studying Drub Abuse*, Rutgers University Press Series in Psychosocial Epidemiology, A. E. Slaby, Series edn., pp. 173–197 (New Brunswick, University of Rutgers Press).

HELZER, J. E., ROBINS, L. N. & DAVIS, D. H. (1975–1976) Antecedents of narcotic use and addiction: a study of 898 Vietnam veterans, *Journal of Drug and Alcohol Dependence*, 1, pp. 183–190.

HELZER, J. E., ROBINS, L. N. & DAVIS, D. H. (1976) Depressive disorders in Vietnam returnees, *Journal of Nervous and Mental Disorders*, 163, pp. 177–185.

HELZER, J. E., ROBINS, L. N., WISH, E. D. & HESSELBROCK, M. (1979) Depression in Vietnam veterans and civilian controls, *American Journal of Psychiatry*, 136, pp. 526–530.

ROBINS, L. N. (1973) *A Follow-Up of Vietnam Drug Users*, Special Action Office Monograph, Series A, No. 1 (Washington, DC, Executive Office of the President).

ROBINS, L. N. (1974a) *The Vietnam Drug User Returns*, Special Action Office Monograph, Series A, No. 2 (Washington, DC, US Government Printing Office).

ROBINS, L. N. (1974b) A follow-up study of Vietnam veterans' drug use, *Journal of Drug Issues*, 4, pp. 61–63.

ROBINS, L. N. (1975) Drug treatment after return in Vietnam veterans, *Highlights of the 20th Annual Conference* (Perry Point, MD, Veterans Administration Studies in Mental Health and Behavioral Sciences, Central NP Research Laboratory).

ROBINS, L. N. (1977) Estimating addiction rates and locating target populations: how decomposition into stages helps, in: RITTENHOUSE, J. D. (Ed.) *The Epidemiology of Heroin and Other Narcotics*, NIDA Research Monograph Series 16 (DHEW Pub. No. 1 (ADM) 78–559).

ROBINS, L. N. (1978) The interaction of setting and predisposition in explaining novel behavior: drug initiations before, in, and after Vietnam, in: KANDEL, D. (Ed.) *Longitudinal Research in Drug Use: Empirical Findings and Methodological Issues* (Washington, Hemisphere).

ROBINS, L. N., DAVIS, D. H. & GOODWIN, D. W. (1974a) Drug use by US Army enlisted men in Vietnam: a follow-up on their return home, *American Journal of Epidemiology*, 99, pp. 235–249.

ROBINS, L. N., DAVIS, D. H. & NURCO, D. N. (1974b) How permanent was Vietnam drug addiction? *American Journal of Public Health*, Suppl. 64, pp. 38–43.

ROBINS, L. N., HELZER, J. E. & DAVIS, D. H. (1975) Narcotic use in Southeast Asia and afterward: an interview study of 898 Vietnam returnees, *Archives of General Psychiatry*, 32, pp. 955–961.

ROBINS, L. N., HESSELBROCK, M., WISH, E. & HELZER, J. E. (1978) Polydrug and alcohol use by veterans and nonveterans, in: SMITH, D. E., ANDERSON, S. M., BUXTON, M., GOTTLIEB, N., HARVEY, W. & CHUNG, T. (Eds) *A Multicultural View of Drug Abuse* (Cambridge, Schenckman).

ROBINS, L. N., HELZER, J. E., HESSELBROCK, M. & WISH, E. (1980) Vietnam veterans three years after Vietnam: how our study changed our view of heroin, in: BRILL, L. & WINICK, C. (Eds) *Yearbook of Substance User and Abuse* (New York, Human Science Press).

WISH, E. D., ROBINS, L. N., HESSELBROCK, M. & HELZER, J. E. (1979) The course of alcohol problems in Vietnam veterans, in: GALANTER, M. (Ed.) *Currents in Alcoholism*, Vol. VI (New York, Grune & Stratton).

References to other studies

BRUNSWICK, A. (1979) Black youths and drug-use behavior, in: BESCHNER, G. M. & FRIEDMAN, A. S. (Eds) *Youth Drug Abuse*, pp. 433–492 (Lexington, MA, Heath).

KANDEL, D. (1975) Stages in adolescent involvement in drug use, *Science*, 190, pp. 912–914.

HYMAN, S. E. (1992) Addiction: what science knows, *The Harvard Mahoney Neuroscience Institute Letter*, 1, pp. 1–3.

O'DONNELL, J. A., VOSS, H. L., CLAYTON, R. R., SLATIN, G. T. & ROOM, R. G. W. (1976) *Young Men and Drugs: A Nationwide Study*, Research Monograph Series, Vol. 5 (Rockville, MD, National Institute on Drug Abuse).

ROBINS, L. (1966) *Deviant Children Grown Up: A Sociological and Psychiatric Study of Sociopathic Personality* (Baltimore, Williams & Wilkins), Reprinted 1974 as a Behavioral Science Club Book selection (Huntington, NY, Robert E. Krieger).

ROBINS, L. & MURPHY, G. E. (1967) Drug use in a normal population of young Negro men, *American Journal of Public Health*, 57, pp. 1580–1596.

ROBINS, L. N., LOCKE, B. & REGIER, D. (1991) An overview of psychiatric disorders in America, in ROBINS, L. & REGIER, D. (Eds) *Psychiatric Disorders in America* (New York, The Free Press).

STEPHENS, R. & COTTRELL, E. (1972) A follow-up of 200 narcotic addicts committed for treatment under the Narcotic Addict Rehabilitation Act (NARA), *British Journal of Addiction*, 67, pp. 45–53.

WALDORF, D. & BIERNACKI, P. (1979) Natural recovery from opiate addiction: a progress report, Unpublished manuscript.

WIKLER, A. (1973) Dynamics of drug dependence: implications of a conditioning theory for research and treatment, *Archives of General Psychiatry*, 28, pp. 611–616.

WINICK, C. (1962) Maturing out of narcotic addiction, *Bulletin on Narcotics*, 14, pp. 1–7.

ZINBERG, N. E. & JACOBSON, R. C. (1976) The natural history of chipping, *American Journal of Psychiatry*, 133, pp. 37–40.

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.

Poo et al. BMC Biology
DOI 10.1186/s12915-016-0261-6

BMC Biology

FORUM Open Access

What is memory? The present state of the engram

Mu-ming Poo[1*], Michele Pignatelli[2], Tomás J. Ryan[2,3], Susumu Tonegawa[2,3*], Tobias Bonhoeffer[4], Kelsey C. Martin[5], Andrii Rudenko[6], Li-Huei Tsai[6], Richard W. Tsien[7], Gord Fishell[7], Caitlin Mullins[7], J. Tiago Gonçalves[8], Matthew Shtrahman[8], Stephen T. Johnston[8], Fred H. Gage[8], Yang Dan[9], John Long[7], György Buzsáki[7] and Charles Stevens[8]

Abstract

The mechanism of memory remains one of the great unsolved problems of biology. Grappling with the question more than a hundred years ago, the German zoologist Richard Semon formulated the concept of the engram, lasting connections in the brain that result from simultaneous "excitations", whose precise physical nature and consequences were out of reach of the biology of his day. Neuroscientists now have the knowledge and tools to tackle this question, however, and this Forum brings together leading contemporary views on the mechanisms of memory and what the engram means today.

The cellular basis of memory

Mu-ming Poo

Neurobiological studies of memory over the past century have progressed along two relatively independent lines of inquiry: the top-down approach examines the animal's behaviors associated with memory acquisition, consolidation, and retrieval, as well as the brain regions underlying these processes, whereas the bottom-up approach explores the cellular and circuit mechanisms of memory encoding and storage by examining the patterns of neuronal firing and the efficacy of synaptic transmission. In his monumental treatise [1] *The Organization of Behavior* (1949), Donald Hebb made a bold attempt to link these two lines of inquiry by postulating that perceptual memory resides in specific "cell assemblies" formed by the strengthening of interneuronal connections due to correlated activities during memory acquisition. The discovery of activity-induced long-term potentiation (LTP) and long-term depression (LTD) of central synapses in the 1970s and 80s further sparked the interest of a whole generation of neurobiologists in studying synaptic plasticity and its relationship to memory. There is now general consensus that persistent modification of the synaptic strength via LTP and LTD of pre-existing connections represents a primary mechanism for the formation of memory engrams. In addition, LTP and LTD could also lead to the formation of new and elimination of old synapses and thus changes in structural connectivity in the brain. Indeed, early development of neural circuits, whereby neural activity sculpts synaptic connectivity [2], depends on processes similar to that associated with LTP and LTD in the adult brain and could be considered as the imprinting of memory engrams generated by early experience.

In this Forum, a group of experts on the cellular mechanisms of memory were invited to present their views on "what is memory", including where and how memory engrams are stored, consolidated, and retrieved. Drawing on an elegant set of studies, Michele Pignatelli, Tomás Ryan, and Susumu Tonegawa illustrate how recently developed techniques to tag and manipulate neurons have begun to establish a causal link between neuronal activity, persistent synaptic changes, and an animal's memory-associated behaviors. The theme of persistent synaptic changes and their causal role in memory is taken up by Tobias Bonhoeffer, who summarizes the evidence that dendritic spines, where excitatory synapses are located, represent the basic cellular unit for memory; long-term memory is stored in a set of spines that are formed or modified during learning and these changes may persist throughout the animal's life.

Based on the findings of activity-induced transcriptional activation and synapse-specific local translation of proteins, Kelsey Martin expands on the idea that the basic building block of memory is the synapse, where

* Correspondence: mpoo@ion.ac.cn;
[1]Institute of Neuroscience, Chinese Academy of Sciences, Shanghai, China
[2]RIKEN-MIT Center for Neural Circuit Genetics at the Picower Institute for Learning and Memory, Department of Biology and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Full list of author information is available at the end of the article

BioMed Central

© 2016 Poo et al. **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

both pre- and postsynaptic elements together with associated glial processes form an integral unit with an individual identity and distinct "neighborhood". Andrii Rudenko and Li-Huei Tsai redirect attention to the nuclei of engram cells, discussing the evidence that epigenetic alterations of the neurons activated during memory acquisition may be involved in the long-term retention of memory. They propose that such epigenetic modification represents a priming event during the initial phase of memory formation; memory retrieval would then trigger the expression of the primed genes, leading to protein synthesis and synaptic modification at individual synaptic units.

Depending on the availability of cellular resources, immediate modifications (LTP and LTD) and long-term turnover (formation and elimination) of individual synaptic units are bound to influence other units on the same postsynaptic cell. Richard Tsien, Gord Fishell, and Caitlin Mullins focus on the important issue of lateral synaptic interaction and redistribution of synaptic strength associated with LTP and LTD, from the point of view of cellular homeostasis as well as the normalization and signal-to-noise ratio of neuronal activities, and propose a conceptual scheme to address the underlying mechanisms.

The hippocampus is unique in being a key brain region for memory formation and a region in which adult neurogenesis occurs. Associated with hippocampus-dependent spatial memory, Tiago Gonçalves, Matthew Shtrahman, Stephen Johnston, and Fred Gage discuss an intriguing new dimension in the cellular mechanisms of memory formation, whereby continuous addition of newborn dentate gyrus neurons in the adult hippocampus, with their enhanced synaptic plasticity, may contribute significantly to establishing the engram for spatial memory.

As proposed by David Marr in his model of hippocampus-dependent memory [3] and supported by many experimental and clinical studies, episodic memories are transferred after acquisition from the hippocampus to the neocortex for long-term storage. The mechanisms underlying the transfer and consolidation of spatial memory are discussed by John Long and György Buzsaki in the context of hippocampal and entorhinal sharp wave-ripples. These activity patterns occur during sleep or non-attentive brain states and are replays of neuronal firing sequences triggered by recent experience, for example they can be temporally compressed replayed versions of the sequential neuronal firing seen as the animal traverses through a particular environment. As discussed by myself and Yang Dan, although spike timing-dependent plasticity could offer a synaptic mechanism for storing sequence information with intervals up to a few hundreds of milliseconds, it remains largely unknown how neural circuits store and recall the temporal sequence of information up to seconds and longer, periods often associated with episodic memory. A compression of the temporal sequence of events such as occurs during sharp wave-ripples in the hippocampus and neocortex offers a potential solution.

The contributing articles of this Forum reflect the tremendous progress made in our understanding of the cellular building blocks of memory. There is a clear consensus on where the memory engram is stored—specific assemblies of synapses activated or formed during memory acquisition—and a substantial body of knowledge on how the engram is generated and maintained in the brain. However, knowing the building blocks and their properties is far from understanding the architecture of the "memory palace". As Charles Stevens indicates in his epilogue, and the readers will soon discover, many new territories are now open for exploration.

Engram cell connectivity as a substrate for memory storage

Michele Pignatelli, Tomás J. Ryan, and Susumu Tonegawa

The storage of information refers to the systematic process of collecting and cataloging data so that they can be retrieved on request.

One of the most enlightening conceptualizations of the neural representation of stored memory information was developed by Richard Semon, who conceived the Engram Theory, a theory of memory traces [4]. According to this theory, as fortified by contemporary knowledge, learning activates a small ensemble of brain cells, inducing in these cells persistent physical/chemical changes. In addition, reactivation of these cells by relevant recall cues results in retrieval of the specific memory. The theory poses an important question: what is the nature of the persistent changes?

In his seminal book published in 1949, Donald Hebb proposed a mechanism based on synaptic plasticity as a substrate of memory [1]. With an example of two cells connected by an excitatory synapse, if the activation of one cell leads to the activation of the second one, the connection between the two cells is reinforced, a postulate that has been confirmed experimentally [5–8]. The increase in connectivity strength within a diffuse group of cells in a more complex feedforward circuit results in the emergence of an engram cell ensemble.

The systematic dissection of the molecular mechanisms involved in synaptic plasticity has revealed that the cascade of events underlying the plastic changes requires two distinct phases [9, 10]. In the encoding phase, also known as early long-term potentiation (E-LTP), an increase in intracellular Ca^{2+} concentration mediated by post-synaptic NMDA receptors elicits a change in synaptic weight by increasing the insertion and

the conductance of AMPA receptors [11]. The dendritic spines that support the post-synaptic machinery rapidly increase in number [12]. In a second phase lasting a few hours after the initial encoding period, the increased synaptic weight is maintained by a protein synthesis-dependent process known as cellular consolidation, during which the steady state synthesis of AMPA receptors is shifted to a higher level. This second phase is known as late LTP (L-LTP) [9, 10] and is sensitive to protein synthesis inhibitors (PSI).

To date, memory storage has mainly been investigated by pharmacological or molecular manipulation and by correlating synaptic changes with the strength of memory recall. Only recently has it become possible to specifically tag cells activated by learning. The demonstration that these cells are part of a memory engram ensemble was provided by a series of "optogenetic" experiments where a learning-induced tagging strategy was employed to express channelrhodopsin [13] in a small population of learning-activated cells [14, 15]. The opsin allowed for the artificial light-induced reactivation of the cellular population labeled during learning and resulted in memory retrieval. The same tagging strategy also verified the reactivation of tagged cells upon presentation of retrieval cues.

Tagging "engram" cells offers a straightforward opportunity to investigate the nature of the persistent changes that occurred in these cells in response to learning. In a recent study [16], "engram cells" were compared to "non engram cells" (non-tagged cells) by ex vivo patch clamp recordings after contextual fear conditioning (CFC). Engram cells displayed changes in synaptic weight typical of LTP such as high current amplitude, insertion of AMPA receptors, high spontaneous excitatory post-synaptic current frequency and amplitude, and increased dendritic spine density. These changes were blocked by the systemic injection of PSI specifically within the consolidation window. Therefore, it is now clear that cells recruited by learning display synaptic changes typical of LTP and are reactivated by retrieval cues, and their reactivation can elicit memory recall. Remarkably, however, protein synthesis-dependent L-LTP seems to be dispensable for memory storage because direct optogenetic activation of the engram cells in PSI-injected mice elicited full memory recall under a variety of conditions.

If L-LTP is dispensable for memory storage, what mechanism is responsible?

An integral memory engram may consist of preferential connectivity between engram cell ensembles distributed across multiple brain regions. In the same report [16] it was shown ex vivo that engram cells from the dentate gyrus established preferential connections with engram cells in the downstream hippocampal CA3 region in a feedforward excitatory engram cell circuit. Remarkably, this preferential connectivity was maintained in mice rendered amnesic by treatment with PSI within the consolidation window, suggesting that memory storage may survive retrograde amnesia in the form of a neural connectivity pattern. Indeed, optogenetic stimulation of DG engram cells in vivo elicited similar cellular reactivation patterns not only in the CA3 region but also in the amygdala for both control and amnesic groups, thus confirming the persistence of engram cell connectivity. These observations support the concept of preferential connectivity of engram cell ensembles distributed across multiple brain regions, which is established during learning and persists despite disruption of consolidation and thereby provides a lasting substrate for memory storage. These data also suggest that the synaptic potentiation observed in consolidated engram cells is necessary for memory retrievability and not for storage [17]. While regulation of synaptic weight provides a scalar quantity to control information retrieval, synaptic connectivity holds the information specificity. This is because synapses that are activated during the encoding stage will dictate the eventual pattern of cellular connectivity of the upstream and downstream engram cell ensembles. This notion is compatible with the broad view that synapses are the basic units of information storage (see Bonhoeffer, this Forum).

How would the specific connectivity pattern between engram cell ensembles be formed by specific learning?

Neural connections are formed during development and certain circuits hold the innate capability to elicit complex behavioral reactions in response to specific perceptual cues [18]. However, this does not seem to be the case in the hippocampal formation because inactivation of the downstream CA1 region before CFC results in anterograde amnesia which cannot be bypassed by direct optogenetic stimulation of DG engram cells [16]. Thus, the memory circuit is not configured under anterograde amnesic conditions and is not, therefore, genetically determined but requires hippocampal activity during memory encoding. As reported by Ryan et al. [16], learning-induced changes in connectivity patterns are insensitive to PSI. So is E-LTP [9, 10] and this early phase of plasticity may provide a framework to investigate the formation of new connections. For instance, blocking NMDA receptor function should impair the emergence of learning-induced connectivity patterns.

During the induction of LTP, existing connections can be potentiated [19] but new connections can also emerge [20]. A hypothetical way this might happen is through the activation of silent synaptic connections [21]. These synapses expressing only NMDA receptors and not AMPA receptors can become unsilenced through AMPA receptor insertion, a mechanism that could, in principle,

support the formation of learning-induced changes in engram cell connectivity. Another possibility is that local dendritic protein synthesis contributes to the rapid synapse formation on engram cells, independently of its role in synaptic potentiation (see Martin, this Forum).

What then maintains the learning-induced engram cell connectivity that is initiated during encoding?

Although learning-induced synaptic potentiation would be suppressed by PSI, the new connectivity pattern could persist through unsilenced synaptic connections of basal unpotentiated strength (Fig. 1). Consistent with this perspective, it has been recently shown that optogenetically induced long-term depression (LTD) of amygdala cells impaired existing conditioned fear responses but subsequent optogenetically induced LTP of the same cells could restore optogenetic cue-evoked recall of the fear memory [22].

The ability to tag cells activated by learning has opened up new horizons in the investigation of memory

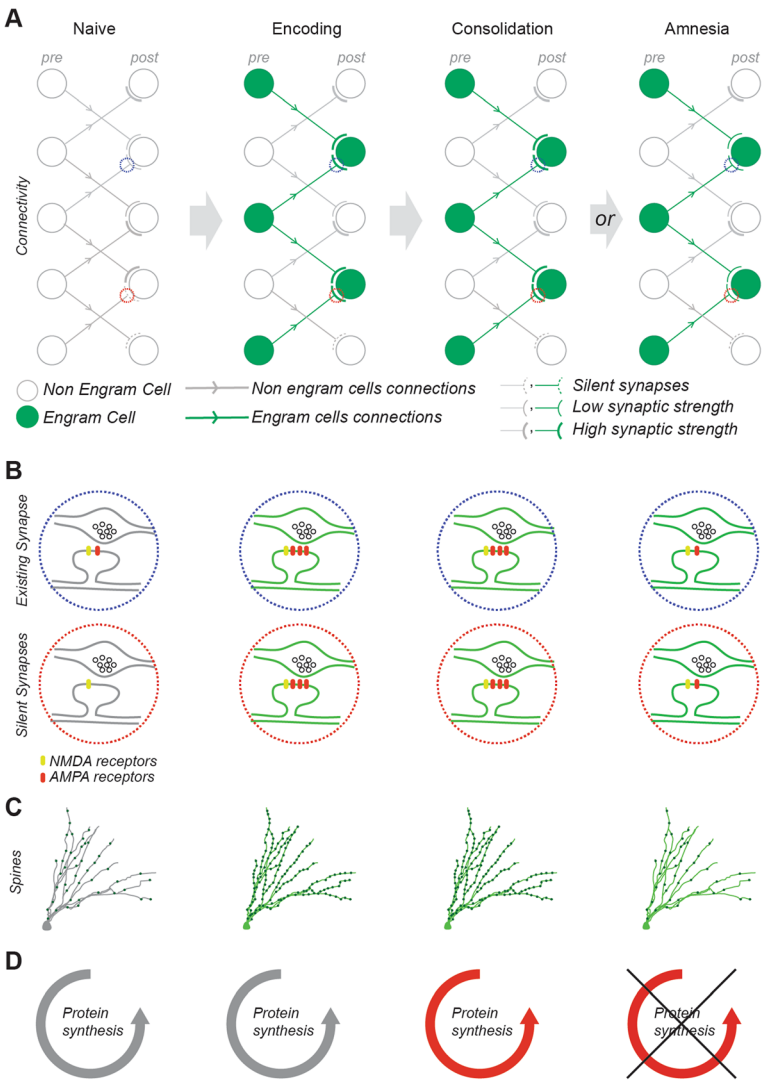


Fig. 1. Synaptic connectivity between engram cells as a mechanism for memory storage. **a** Cellular connectivity in a feedforward excitatory circuit, **b** synaptic configuration, **c** dendritic spine density, and **d** protein synthesis state, shown in a naïve circuit, a circuit during encoding, a circuit after consolidation, or a circuit in an amnesic condition. Engram circuit, cells, and synapses are displayed in *green*, non-engram in *gray*. In the naïve state, the circuit displays a variety of synaptic patterns, including strong (*thick gray lines*) and weak synapses (*thin gray lines*) as well as silent synapses (*dotted lines*) exclusively expressing NMDA receptors. During encoding, a network of engram cells is recruited. The preferential connection between engram cells occurs either by potentiation of existing connections (*blue dotted circles*) or by unsilencing synapses (*red dotted circles*). A spine density increase supports the synaptic changes. During consolidation, the steady state synthesis of AMPA receptors is shifted to a higher level and the disruption of consolidation with protein synthesis inhibitors (PSI) results in retrograde amnesia. However, during PSI-induced amnesia, memory storage persists within an engram-specific set of weak synaptic connections

by revealing the role of the engram-specific connectivity in the storage of information.

Spines and synapses as basic elements of memory storage

Tobias Bonhoeffer

In his seminal work, Donald Hebb [1] proposed that the basic mechanism by which memories are stored in the brain is the enhancement of synaptic strength and, in connection with that, morphological changes of the respective synaptic contacts. In other words he proposed that synapses and not cells are the basic building blocks of memory, from a theoretical perspective a reasonable suggestion as there are approximately 10,000–100,000 times more synapses in the brain than neurons.

By now it is well established that morphological changes at the synaptic level occur in conjunction with stimuli that are thought to mimic learning events. In vitro experiments [12, 23] have shown that long-term potentiation results in the addition of dendritic spines, tiny protrusions which harbor synaptic contacts. Tens of thousands of these spines decorate the dendrites of most excitatory cells in the hippocampus and the neocortex. And indeed, further studies showed that spines not only come and go but also change their shape during putative learning events [19], a suggestion that had been put forward in a purely theoretical paper by Francis Crick [24]. So, it is well established that spines emerge, disappear, and change with cellular events thought to underlie learning processes. But is this merely a correlation or are there ways towards showing that these events really lie at the basis of learning and memory storage in the brain? Recent experiments have made substantial progress in this respect.

The first study that made a clear case that spines are important for the long-term storage of information was done in the visual cortex of mice [25]. In the visual cortex it is well known that synaptic connections are established or modified with changes in visual experience, like the temporary closure of one eye. These plastic changes are often used as a proxy for what happens during memory formation since they share key features: it is, for instance, a universally accepted fact in memory research that information that has been acquired early in life can be learned much more easily a second time, even if it had been "completely forgotten" in the meantime. This effect has been called "savings" [26] and is a hallmark of most memory processes. It has been shown that the same effect occurs in the visual system [27]. Mice were monocularly deprived for a couple of days early in life so that the visual system adapted to this change of the visual environment. Subsequently, animals were subjected to normal vision again so that their visual cortex reverted to normal function. If monocular deprivation was then performed a second time, much later in life, when normally this procedure has only a very limited effect (if any), substantial adaptation still takes place because of the early experience that the animal has had. Importantly, this savings effect could be related to new spines that emerged during the first plasticity episode and persisted [25]. The fact that there was no growth of additional spines during the second plasticity period, while the functional adaptation occurred much faster and more reliably, suggests that the persistent spines facilitate the second adaptation [25]. Therefore, these spines serve to "remember" the previous sensory experiences the animals had. Two subsequent studies [28, 29] further bolstered the case by showing that also in the motor cortex the generation of new spines forms the basis of learning motor tasks of different sorts. Interestingly, in one of these studies [28] it was also found that relearning a task occurred faster and did not involve the generation of new spines, again arguing for persistent spines "memorizing" specific motor tasks. Furthermore, this study demonstrated that learning different tasks involves different sets of spines, providing a strong argument for spines and not cells being the relevant entity for information storage in the brain.

These three papers were among the first to make a strong case for a causal relationship between new (or changing) spines and learning or information storage in the brain. Subsequently, a number of studies further strengthened this hypothesis. Some of them used fear conditioning to show that also in this paradigm learning is paralleled by structural changes: fear extinction and fear conditioning are marked by the generation or removal of spines in the frontal association cortex [30] and the auditory cortex [31]. One particularly interesting finding in this context is that extinction induces appearance of spines that were eliminated upon the original fear conditioning to the same stimulus but not to a distinct conditioned stimulus, suggesting that the spines are again specifically associated with extinction of one specific association [30]. Interestingly, also in a completely different animal model—song learning in zebra finches—it was shown that new spines are generated in the forebrain nucleus HVC when an animal learns a new song from a tutor [32].

Finally, what about the experiment that has long been on the agenda [33], namely to specifically ablate spines that have been generated during learning? If the above interpretations are true, spine ablation should lead to forgetting of the information that was learned when the new spines were generated. First important strides in that direction have been made in a recent experiment by the group of Haruo Kasai [34], who specifically labeled spines that were generated at a particular time window immediately after learning. When these spines were later

ablated or at least reduced in size, the animal indeed forgot the previously learned information. The learning paradigm is so far relatively simple (rotarod learning) but it provides a very nice indication that the generation of new spines or their enlargement is truly causal at least in some forms of learning.

Taken together, there is now considerable evidence from different species as well as from different learning paradigms that spines, and thus synapses, change when an animal learns. Furthermore, there are convincing indications that the maintenance of previously established structural connections on the level of dendritic spines explains the memory phenomenon of savings. Finally, if spines are ablated, an animal forgets what it has learned through the addition of new or stronger spine synapses. All of these experiments seem to point strongly towards the notion that spines or synapses (and not entire cells) may be the smallest unit of memory storage in the brain and it may, therefore, be most appropriate to say that the "engram" of a memory is laid down in the set of spines or synapses that are changed when specific information is stored. This is of course not to say that engrams are not visible on the level of single cells (see preceding contribution to this Forum by Pignatelli, Ryan, and Tonegawa); after all, the activity of cells is determined by the complement of their synapses. Yet, the finest resolution of the engram may only become apparent if one truly considers everything on the basis of the pattern of synapses or spines which are changed during a particular memory event.

A cell biologist's view of memory: revisiting the neuron doctrine

Kelsey C. Martin

Over a century ago, the anatomist Ramon y Cajal used the Golgi staining method to visualize individual cells in the brain. He observed that the brain was composed of discrete cells rather than of a "reticular network" of interconnected cells (as was commonly believed at the time). Cajal's observations provided critical support for the neuron doctrine, which postulates that the neuron is the central unit of the brain. At the same time, his drawings revealed the beautiful complexity, polarity, and compartmentalization of neurons: cell bodies elaborating axonal and dendritic processes, forming up to thousands of synapses with one another. With a remarkable amount of prescience, Ramon y Cajal also speculated that memories were stored as increases in the numbers of connections between neurons. This idea forms the basis of learning-related synaptic plasticity—the idea that memories are stored as changes in the number and strength of synapses between neurons—a framework that has endured as a model for understanding the biology of memory.

Studies of memory in rodents and goldfish performed in the 1960s and 1970s demonstrated that long-term memory required protein synthesis whereas short-term memory did not [35]. Subsequent studies of learning-related plasticity in organisms ranging from *Aplysia* sensory-motor synapses to rodent hippocampal synapses similarly demonstrated that long-lasting forms of plasticity can be differentiated from short-term plasticity by their dependence on RNA and protein synthesis [36]. Molecular biological approaches led to the identification of many genes that contribute to long-term plasticity and memory and to the elucidation of specific patterns of neuronal activity and specific signaling pathways that trigger changes in RNA and protein synthesis within neurons. These findings gave rise to the idea that activity-dependent changes in the neuronal transcriptome and proteome mediate and/or maintain the changes in neuronal structure and physiology that result in persistent changes in synaptic strength.

Initially, studies of gene expression underlying memory focused on activity-dependent changes in transcription in the nucleus. The discovery of immediate early genes, such as c-fos, arc, and zif268, that were induced during memory formation implied that the neuron was the unit of long-term plasticity and memory [37]. Immediate early genes are now widely used to map neurons involved in memory formation, in line with the idea that memories are encoded within networks of discrete neurons in the brain.

In contrast to this neuron-centric view of memory formation, studies of learning-related plasticity revealed that plasticity was synapse-specific, that is, it could occur at some but not all synapses made by an individual neuron [38]. This finding raised questions about how the products of gene expression, synthesized in the nucleus, could be targeted to alter structure and function at subsets of synapses within a single, highly polarized and compartmentalized neuron. One solution to this question came from findings that mRNAs localized to distal dendrites and synapses [39], where their translation could be regulated by activity. These findings included detection of polyribosomes at the base of synapses as well as the identification of a subset of mRNAs that localized to distal dendrites. Studies from several labs have identified hundreds to thousands of dendritically localized mRNAs [40–42] translation of which could alter the structure and function of synapses, thereby making the synapse (or neighborhood of synapses), rather than the entire neuron, the unit of plasticity.

Local translation at synapses has been shown to regulate translation in a synapse-specific manner [43, 44] and inhibiting local translation at synapses in a variety of in vitro preparations has been shown to block long-term learning-related plasticity [45, 46]. In one of these

preparations, the *Aplysia* sensory-motor culture system, we directly tested the relationship between transcriptional regulation in the nucleus and translational regulation at the synapse during synaptic plasticity and found that stimulus-induced newly transcribed localized mRNAs were delivered throughout the neuron but were only translated at locally stimulated synapses [47]. The implication of this finding is that activity-dependent transcription sets the entire neuronal arbor in a state of readiness to respond to local cues via regulated translation of localized mRNAs. From the perspective of activity-dependent gene regulation, this idea shifts the focus from the neuron as the unit of plasticity to the synapse (or neighborhoods of synapses). In so doing, it underscores the need to refine the current neuron-based maps of memories by developing synaptic maps of memory in the brain (Fig. 2).

The focus on the synapse as the unit of plasticity calls into question aspects of the neuron doctrine. Thus, while the neuron doctrine assumes that neurons are separate, autonomous units, cell biological studies of synapses indicate that all components of the synapse are intimately interconnected, including not only the pre- and post-synaptic compartments but also glial processes. For example, in one study of local translation at *Aplysia* sensory-motor synapses, we found stimulus-induced local translation of a reporter depended on trans-synaptic signals from the post-synaptic motor neuron compartment to the presynaptic sensory neuron compartment [44]. Taking the idea of non-cell autonomy to an extreme, recent studies have reported that neurons and/or glia can transfer mRNAs and microRNAs to neighboring neurons and/or glia [48]. This finding raises the provocative idea that gene expression can be regulated by transfer of genetic information from cell to cell. As such, the idea of "local" translation serving to alter the structure and function of synapses encoding a memory suggests that "synaptic maps" of memory are not encoded by discrete synaptic compartments of individual cells but rather by complex fluid synaptic neighborhoods. These neighborhoods include not only the products of gene expression that are made in the nucleus of the neuron the synapse belongs to but also signals (including RNAs) from neighboring neurons and glia.

To add an additional layer of complexity onto this view of memory, single cell RNA sequencing experiments have uncovered remarkable complexity in the nerve cells and glia within the brain [49]. These findings indicate that synaptic neighborhoods are not just comprised of common modules of pre-synaptic, post-synaptic, and glial components but that they also exist in a multitude of distinct flavors or "ethnicities" as neighborhoods. A recent electron microscopic reconstruction of a small volume of mouse neocortex revealed that the formation of synapses between neurons is more dependent on cell identity than on proximity [50]. The diversity of cell identities, in combination with the idea that cell identity drives synapse formation, indicates that generating a synaptic map of memory will require understanding the distinct identities of the participating cells as well as the details of their cell-to-cell interactions.

As a cell biologist interested in understanding memory, the challenges moving forward include identifying conserved local processes that persistently alter synaptic function and developing methods to manipulate these processes in order to test their function in the formation and storage of memory. Possibilities include local mechanisms of translational regulation at synapses, trans-

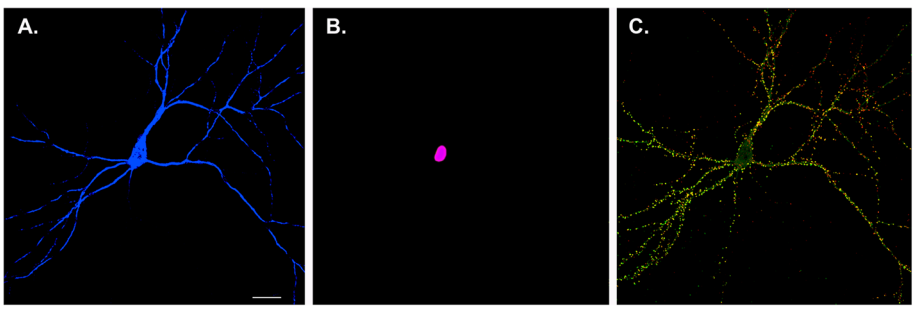


Fig. 2. Neuron versus synaptic maps of memory. The Golgi method used by Ramon y Cajal at the turn of the 19th century revealed that the brain consisted of individual nerve cells, leading to the formulation of the neuron doctrine. In modern neurobiology, expression of soluble fluorescent proteins allows visualization of individual neurons in the brain or, as shown here, in a hippocampal pyramidal neuron in dissociated culture (*blue*) (**a**). Identification of immediate early genes, such as cFos, provides a means of labeling the nuclei of individual neurons that are activated to undergo transcription following neuronal activity (*violet*) (**b**). This provides a neuron-level map of activity. However, each neuron forms thousands of synapses, shown in **c** by labeling a single neuron with presynaptic (*green*) and postsynaptic (*red*) markers (*yellow* where the pre- and post-synaptic elements are adjacent to one another). Note that this neuron is connected to other neurons in the culture that are not shown. Developing a synaptic rather than a neuron-wide map of memory requires obtaining activity-dependent markers that label the activated synapses rather than nuclei. Understanding synaptic maps will not only require understanding the identity of individual neurons but also the details of their cell-to-cell interactions

synaptic signaling pathways (including the transfer of RNAs via extracellular vesicles), interactions between synaptic cell-adhesion molecules, and even extracellular matrix dynamics. As an example of the latter, Roger Tsien has proposed that memories may be stored in the pattern of holes formed within the perineuronal net, a specialized extracellular matrix structure that is formed at the end of critical periods in the brain [51]. Taken together, recent lessons learned suggest that elucidation of these local processes requires consideration of the synaptic compartment as a local environment rather than as a collection of separate and autonomous membrane bound compartments. While morphology, from Golgi stains to electron microscopy, emphasizes the boundaries between cells in the brain, molecular cell biological studies uncover a fundamental role for local cell–cell interactions and interconnections in the encoding of memories.

Memory as a functional consequence of epigenetic priming in engram neurons

Andrii Rudenko and Li-Huei Tsai

Memory formation, storage, and recall constitute the essence of human nature. The search for the mechanisms underlying learning and memory has revealed the importance of a number of molecular and cellular processes, such as activity-dependent gene expression, intracellular signaling cascades, and synaptic plasticity [52, 53]. The long-lasting attempts to characterize memory localization recently resulted in identification of specific neuronal populations—so-called engrams—that provide a physical location for the storage and retrieval of memory traces [14, 15, 54, 55]. Despite discovery of the engram cells, molecular mechanisms of memory storage remain unclear. We propose that epigenetic alterations taking place in these cells may represent a critical process involved in the long-term retention of memory traces.

One well-studied example of such alterations is histone acetylation, a covalent mark of active chromatin. Mutations in CBP, a gene product necessary for the acetylation of multiple memory genes, result in severe intellectual disability in humans as well as in mice [56–58]. Conversely, histone deacetylase inhibitors (HDACi) were shown to restore histone acetylation and ameliorate cognitive deficits in a CBP-deficient mouse model [56, 57]. Moreover, HDACi have been found to ameliorate memory deficits in mouse models of Alzheimer's disease [59, 60]. In addition to histone acetylation, several other epigenetic mechanisms, including DNA methylation and hydroximethylation, have also been demonstrated to regulate memory function [61, 62].

While inhibiting HDACs was found to be effective in enhancing synaptic plasticity and memory, such inhibition per se, in the absence of neural stimulation, produced very limited results [63, 64]. Thus, HDAC inhibition appears to convey its effects on learning and memory via facilitating gene expression elicited by neural activity, a phenomenon known as epigenetic priming [65, 66].

As discussed earlier in this Forum by Pignatelli, Ryan, and Tonegawa, a recent study by Ryan et al. [16] has elegantly demonstrated that inhibiting protein synthesis during the memory consolidation window does not disrupt memory retrieval by means of engram activation. This exciting observation suggests that while augmented synaptic strength may be critical for memory encoding, some other mechanisms, potentially involving epigenomic modifications in engram neurons, appear to be necessary for memory trace storage. The recent discovery that long-term memory can be re-instated following erasure of its synaptic expression strongly supports this idea [67]. We propose that, mechanistically, the engram cells are marked, or tagged, not only synaptically [68] but also at the epigenetic level, be it histone acetylation, methylation, DNA methylation, or Topoisomerase IIβ-dependent topological changes of the DNA/chromatin, as recently suggested by our work [62, 69]. Specifically, the initial phase of memory formation would cause changes in the epigenetic state of the engram cells through a priming event which may be protein synthesis-independent (for example, epigenetic modifications to make specific genomic regions poised for efficient transcriptional activation). Such changes may also lead to long-lasting alterations in chromatin structure and function underlying the memory consolidation process. Finally, memory retrieval would signal the engram cells potentiating initial epigenetic priming, including molecular events such as generation of DNA breaks within the promoter areas of early response genes such as *c-Fos*, *Npas4*, *Nr4a1*, and *Egr1* [69], triggering expression of the primed genes leading to protein synthesis and increases in the number and strength of the synapses. Such a chain of events may explain why, even after considerable neurodegeneration, HDAC inhibition coupled with behavioral training is capable of reinstating learning and retrieval of long-term, and even remote, memory [59, 70]. This scenario may be possible if engram cells, epigenetically primed by the initial learning experience and capable of re-engaging in the chain of molecular events leading to memory retrieval, still remain in the brain after neurodegeneration.

We should note that in these early days of functional memory engram investigation, we still do not have satisfactory answers to many important questions. For example, the exact molecular and structural features of engram-containing networks, or potential mechanisms that might allow participation of a specific neuron in different engram ensembles, currently remain unknown. Moreover, although there is accumulating evidence of

the epigenetic marks/tags that would prime engram cells for an efficient transcriptional response, the exact nature of those marks remains unclear. Deciphering such priming signatures will help us immensely in understanding the mechanistic basis of memory processing.

Memory mechanisms: LTP and LTD in partnership, not merely in opposition

Richard W. Tsien, Gord Fishell, and Caitlin Mullins

How memory is stored in the brain can be effectively queried by examining how this process is affected in neuropsychiatric conditions. To this end, many aspects of the emerging pathophysiology of autism spectrum disorders (ASD) may provide insight into the tuning of synaptic strength in memory [71]. In ASD and related intellectual disability (ID), gene discovery points to dysregulation of interrelated neuronal functions, including control of nuclear gene expression, local protein synthesis in dendrites, and excitation:inhibition (E:I) coordination. These functions are linked together in feedback loops involving electrical or chemical sensors, termed "homeostats". The feedback loop may malfunction as a result of disease-causing mutations in any of the components. Given the effects of these disorders on cognition, it is likely no accident that the same set of functions loom large in current thinking about memory. Indeed, consideration of ASD and other neuropsychiatric diseases provides fresh perspective on the basic underpinnings of memory. From this viewpoint, we offer some thoughts about the relationship between LTP and LTD and the way that information in the brain may be stored and retrieved.

In an intriguing study published in 2012, Bourne and Harris used serial section electron microscopy to compare dendrites receiving either theta burst LTP or control stimulation [72]. As expected for LTP [34, 73], they found single spines with increased postsynaptic area by 2 h. Surprisingly, however, this occurred concomitant with a remarkable reduction in the number of small spines, leaving the total postsynaptic area per unit dendritic length unchanged compared with control (Fig. 3a). The constancy of total postsynaptic area per unit dendritic length fits well with prevailing concepts about homeostasis and normalization, but it also raises provocative questions about the neurobiological mechanism and organization of memory.

Two previously proposed and rather different hypotheses considered here might explain the underlying basis

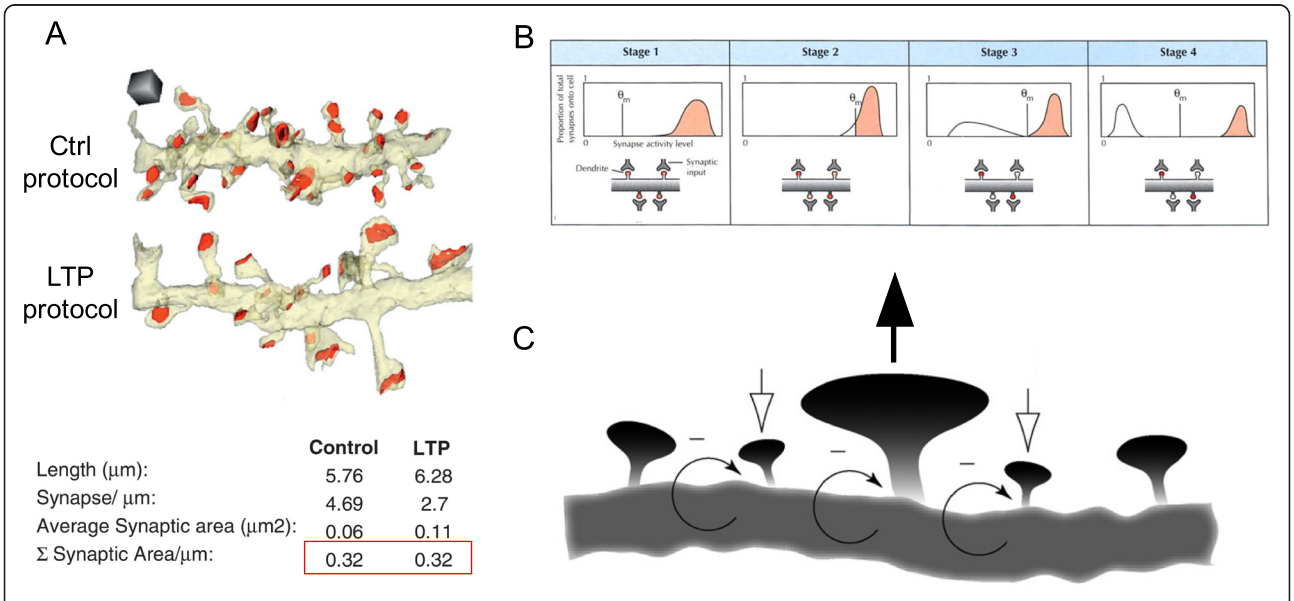


Fig. 3. Possible mechanisms of synaptic modification in memory storage. **a** Structural synaptic scaling is analyzed by serial section electron microscopy in hippocampal tissue under control and LTP (induced via theta burst stimulation) conditions. Control and LTP dendrites have equal postsynaptic areas (*red outline*) despite differences in synaptic density and size. Adapted from [72], with permission. **b** Representation of metaplasticity occurring, shifting the plasticity threshold (θ_m). The graphs on the *top row* depict the distribution of synaptic responses, *red shading* indicating LTP and *white shading* LTD. The four synapses illustrated (*bottom row*) reflect how the shifting of θ_m would impact the strength of individual synapses (intensity of *red shading* correlating to strength). Stage 1 depicts a saturating event, where all of the synaptic strengths have been excited to levels above θ_m. Stage 2 shows a metaplastic response to the activity levels reached in stage 1. In stage 2, θ_m has shifted such that some synapses are below the new threshold and weaken in stage 3. This changed θ_m leads to a new stable state, illustrated in stage 4. Reproduced from [75], with permission. **c** Synaptic potentiation at one spine (center, *upward-pointing black arrow* added for emphasis) is predicted to temporarily increase activity at adjacent spines. In response to the increased activity, homeostatic plasticity weakens the entire area (*negative feedback arrows*), with closely neighboring spines being disproportionately weakened (*downward-pointing white arrows*). The weakened spines allow the overall dendritic length to maintain a constant level of activity with the central spine still maintaining a level of potentiation. Adapted from [76], with permission

of Bourne and Harris' observations. In reviewing evidence for a "sliding threshold" concept [74], synaptic plasticity with flexible rules (metaplasticity), Deisseroth et al. [75] discussed a scheme that could fit with Bourne and Harris' results (Fig. 3b). If a stretch of dendrite receiving strong inputs undergoes LTP but then autoregulates the threshold dividing LTD and LTP, it would attain a new equilibrium that fractionates the synapses into stronger and weaker subpopulations (Fig. 3b). Also pertinent to Bourne and Harris' results, Rabinowitch and Segev [76] focused on lateral coordination of synaptic strength, invoking unknown mechanisms of local regulation on dendritic branches. In their hypothesis, LTP at spines receiving strong synaptic input is yoked together by a compensatory mechanism that weakens, or even eliminates, nearby neighbors (Fig. 3c).

Although these papers make no reference to each other, they converge on a shared view of synaptic plasticity, jointly supported by biological evidence and theoretical rationale: a synapse may undergo Hebbian strengthening or weakening as an individual entity but, over time, it can also behave as a connected entity, operating in coordination with other synapses in the same neuron or dendritic branch. Further, LTP and LTD can cooperate to redistribute synaptic weight. This notion differs from the traditional analogy between synapses and digital information storage devices, in which bits are stored and retrieved independently. On the other hand, coordination amongst multiple synapses, made by different inputs, provides benefits with regard to issues of normalization and signal-to-noise.

What can be said about the molecular mechanism(s) that might allow or even drive the coexistence of LTP and synaptic weakening in close proximity along a stretch of dendrite? This question has received much less attention than strengthening and weakening of the same synapses (see, for example, an elegant demonstration of LTP and LTD in direct opposition [22]). In the spirit of this Forum, we list here multiple possibilities for such lateral interaction:

1. Concentration of the excitatory neurotransmitter glutamate ([Glu]) must fall off with increasing distance from a strong input and could play some role in lateral interactions. Whereas NMDA receptors are essential for most forms of postsynaptically expressed LTP and are driven by high [Glu], lower levels of [Glu] would be sensed by the more sensitive metabotropic glutamate receptors (mGluR) at nearby synapses and could foster LTD mechanisms.
2. Different programs of local protein synthesis may be triggered by NMDARs and mGluRs and support LTP (e.g., increased synthesis of AMPA receptors) or LTD (e.g., upregulation of Arc). As we propose elsewhere [71], mutual inhibition between such translational programs could help enforce a sharp threshold dividing LTP and LTD, whereby local signals mediate LTP proximally and LTD at a distance. Interestingly, defects in either program can give rise to ASD.
3. βCaMKII (beta calcium/calmodulin-dependent protein kinase II) is well-suited to serve as a local sensor of activity-dependent rises in Ca^{2+} because of its high Ca^{2+}/CaM sensitivity. Hence, βCaMKII can also serve as an arbiter, dictating the decision between exo- and endocytosis of AMPARs to promote LTP/LTD. High [Ca^{2+}] activates βCaMKII, increasing AMPAR exocytosis [77]. In contrast, low [Ca^{2+}] leads to CaM-free, kinase-deactivated βCaMKII interacting with Arc, facilitating AMPAR endocytosis [78].
4. AMPA receptors can also be redistributed along the dendritic length by coordination of exocytosis of AMPARs and LTP at one site with endocytosis of AMPARs and LTD in flanking regions. A frank lateral transfer of AMPARs [79] could support a coupling of LTP/LTD at nearby dendritic spines.

Many studies on the cell biology of the dendrite take on different meaning if considered in this light. Also, an organizational principle of "robbing Peter to pay Paul" might engender marked strengthening of some synapses (rich getting richer) at the expense of others (poor getting poorer), thus contributing to empirical observations of a highly skewed distribution of synaptic weights [80–82]. It remains to be seen, however, whether the observations of Bourne and Harris are relevant to behavioral memory in vivo and whether the tradeoff of synaptic strength is predominantly a local, dendritic branch-based phenomenon as suggested by Rabinowitch and Segev [76] or also strongly dependent on regulation at the neuron-wide level [75, 83]. We have spoken little about the temporal dimension but the static snapshots in Fig. 3 are clearly shorthand for complex dendritic dynamics. Whatever the molecular/subcellular mechanism and dynamics, tradeoffs in synaptic strength will have a strong influence on the way that memory traces are stored and retrieved. As the debate continues about memory engrams and persistent changes at the level of synapses, synaptic neighborhoods, ensembles of cells, and even whole circuits, we would do well to consider redistribution in net synaptic strength as an underlying mechanism, rather than merely net increases or decreases. Perhaps we should modify silicon-based notions of memory units with independent read-write capabilities and embrace assemblies—not just "cell assemblies" [1, 17] but also "synaptic assemblies"—in their full spatiotemporal glory.

A role for adult-born neurons in pattern separation

J. Tiago Gonçalves, Matthew Shtrahman, Stephen T. Johnston, and Fred H. Gage

Memory involves the complex interplay between forming representations of novel objects or events and developing generalizations of similar experiences. Distinct instances of similar events must be discriminated—for example, being able to find your car at work despite parking in a different spot every day—but, at the same time, experiencing just a fragment of a familiar experience, such as a particular smell, can trigger specific memories from your childhood. The interplay between forming distinct memories and generalizing events is conceptualized to involve two separate processes: pattern separation and pattern completion.

The dentate gyrus (DG) is the information gateway of the hippocampal formation and as such plays a crucial role in hippocampal function, including the formation of episodic memories. Additionally, the DG is one of only two regions within the mammalian brain that generates new neurons throughout the life span of an individual in both rodents and humans [84–86]. There is increasing evidence that adult-born dentate granule cells (DGCs) are important for fine pattern separation of similar but distinct events [87, 88].

The DG as a pattern separator

The elegantly delineated anatomy of the hippocampus has long served as a substrate for theories about memory formation. It is grossly delineated into a series of interconnected loops within four major subregions. Specifically, the DG of the hippocampus receives excitatory input from the entorhinal cortex (EC) via the perforant path [89]. The DG then projects through the mossy fiber tract to form powerful synaptic connections with CA3 neurons, which in turn project to CA1, forming the classic trisynaptic pathway. In addition, the DG projects to and receives input from local inhibitory interneurons, most notably from the hilus.

In rodents, the DG contains approximately four- to fivefold more neurons than up- or downstream EC and CA3, respectively. Thus, input from relatively few cells is processed by a much larger neural network within the DG before generating a condensed output. However, only a small percentage of DGCs is activated in response to a given event [90, 91]. Based on these characteristics, modeling studies postulate that the DG is a competitive network that can function as a pattern separator by partially de-correlating inputs [3, 92]. One prediction from this theory is that the DG is critical for forming memories of events that are similar but not identical to each other. This prediction is supported by accumulating evidence from both high-resolution functional MRI studies in humans and studies of "behavioral pattern separation" in rodents [93, 94], where subjects discriminate between similar environments or sensory stimuli presented at different times.

What mechanistic role do adult-born DGCs play in behavioral pattern separation?

Although the gross anatomical connectivity of the DG suggests a role in pattern separation, the DG's ability to perform this function appears to be further enhanced via its ability to incorporate new neurons. Adult-born DGCs undergo a lengthy process of morphological and physiological maturation before they fully integrate into the local hippocampal network [95]. Each immature DGC enters a critical period of greater plasticity approximately 4 to 6 weeks after it is born. These immature DGCs exhibit greater excitability [96], receive less inhibition from local interneurons [97], are more broadly tuned to input stimuli [98], and exhibit greater synaptic plasticity [99] than mature cells. Therefore, immature adult-born DGCs may perform unique computational tasks critical for hippocampal function, in particular behavioral pattern separation [87, 100].

The physiological properties of immature neurons at the single-cell level, however, might yield counterintuitive predictions at the circuit level. While hyperexcitability and broad tuning of immature neurons may allow the hippocampus to encode novel stimuli, their broad tuning would also seem to suggest that a hippocampal circuit rich in immature neurons would fire more frequently and indiscriminately to various environmental inputs. Yet, this explanation appears to conflict with the role of the DG in pattern separation. Therefore, it has recently been proposed that, though hyperexcitable themselves, these immature neurons' key contributions could be to suppress overall DG activity, maintain network sparseness, and thereby decrease interference between memory representations [101–103] as illustrated in Fig. 4. Individual immature neurons may be more excitable and broadly tuned but it is hypothesized that these cells in turn activate inhibitory interneurons in the DG and hilus, resulting in greater inhibition of mature DGCs. Developing a unified picture of how single-cell physiological features combine to determine circuit-level responses and behavior presents a great challenge for future work.

Future steps for understanding the role of adult-born neurons in the DG

Previous efforts to link DG function to cellular and molecular mechanisms have been hindered by technical limitations. First, although extracellular recording of DGCs has been used to monitor the dynamics of neuronal firing patterns, sparse activity within the DG makes

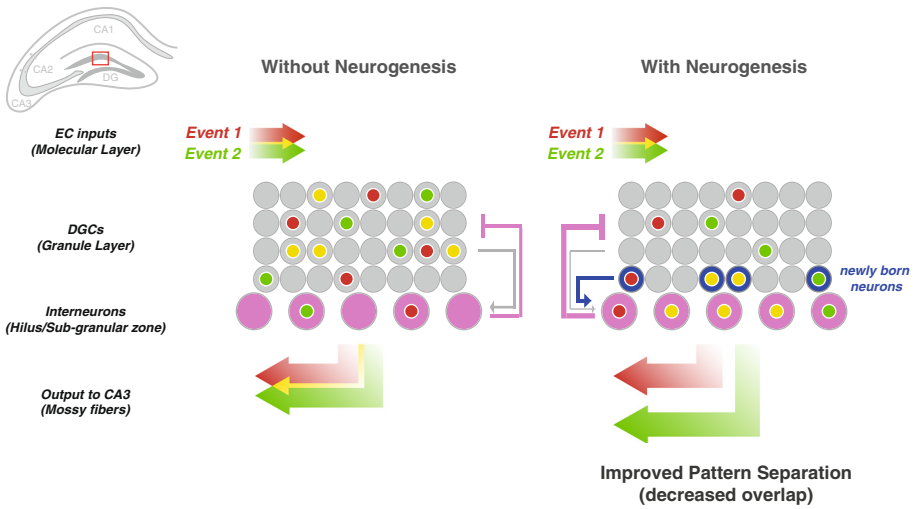


Fig. 4. Immature adult-born neurons improve pattern-separation in the DG by enhancing feedback inhibition. Two events are encoded in separate but partially overlapping populations of activated DGCs in the DG (*red* and *green*, with overlap in *yellow*). DGCs receive strong inhibitory inputs from interneurons (*purple*) in the hilus and sub-granular zone. It is hypothesized that hyperactive immature adult-born DGCs (*blue*) drive these interneurons, enhancing feedback inhibition from the hilus, which results in decreased overlap of activated DGCs and output to CA3, thereby improving pattern separation

it difficult to locate neurons, identify neuronal subtypes, and monitor the activity of a large population of DGCs [104]. Second, immunohistochemical analysis of immediate early gene expression, such as c-fos, can assess activity across a large population but only at limited time points, providing a limited snapshot of DG activity. Nevertheless, advances in in vivo recording techniques promise to eventually enable the monitoring of DG cells during behaviorally relevant tasks. In vivo imaging methods [105] in particular would be perfectly suited for recording DG activity—hundreds of cells can be monitored simultaneously and cellular subtypes identified through the use of genetic markers. Combined with modern techniques to manipulate activity in adult-born neurons, imaging large populations of neurons in vivo would permit testing the hypothesis that immature and mature DGCs work cooperatively to maintain the normal sparse network activity of the DG. These methods also have the potential to provide further insights into the mechanisms behind memory formation and recall, revolutionizing our understanding of the features encoded by both mature and immature DGCs and of how responsiveness to those features changes with learning.

Mechanisms for memorizing temporal sequence and interval

Mu-ming Poo and Yang Dan

The temporal sequence and interval of events are essential elements of episodic and procedural memories. Where and how sequence and interval information is stored in the brain remains a mystery. If modifications of synaptic connectivity are the cellular substrates for memory storage, then the challenge is to understand how distributed synaptic changes within neural circuits represent the sequence and interval of previously experienced sensory or motor events.

In his physiological postulate for perceptual memory, Hebb [1] proposed that reverberating activity generated by perceptual experience along distinct neuronal pathways could provide temporary storage of the experience. Repeated firing of a specific assembly of neurons in a particular temporal sequence could then serve to imprint the memory by modifying synaptic connections among the cells and re-activation of the assembly in this sequence represents recall of the memory.

The popular version of Hebb's postulate refers to synaptic modifications based on correlated pre- and postsynaptic firing ("cells that fire together wire together"), and it received experimental validation with the discovery of long-term potentiation (LTP) [5] and long-term depression (LTD) [106] induced by high- and low-frequency presynaptic stimulation, respectively—high-frequency stimulation results in postsynaptic spiking and thus LTP due to correlated activity, whereas low-frequency stimulation fails to do so, leading to LTD. The discovery of spike timing-dependent LTP and LTD [107, 108] led to further revision of Hebb's learning rule—the sequence of pre- and postsynaptic spiking, rather than simple coincidence of activity, is critical for determining whether the synapse is strengthened or weakened. The defined time windows for spike timing-dependent plasticity (STDP) [8, 109, 110] and its presence at many excitatory synapses [111] suggest that temporal information may be stored via STDP, and Hebb's assembly could be established by sequence-dependent synaptic

strengthening or weakening. In addition to the well-studied sequence replay in the hippocampus (see contribution by Long and Buzsáki in this Forum), this hypothesis was supported by a finding that repetitive visual stimulation with a unidirectional moving spot that evokes sequential spiking of the cortical neurons enhances their sequential firing in response to a flashed stimulus, in a manner that depends on the speed of the conditioning stimulus and activation of NMDA receptors in the visual cortex [112].

The time window for STDP, normally quite narrow (~20 ms) for LTP and wider and more variable (~20–100 ms) for LTD, also imposes a limit on the time intervals between pre- and postsynaptic spiking that can leave a long-term imprint at the synapse. Could the interval be extended via polysynaptic excitation within a network? A simple test of this idea was performed in a random synaptic network formed by dissociated hippocampal neurons in culture. Repetitive pair-pulse stimulation at a fixed interval was applied to a single input neuron and changes in the efficacy of polysynaptic connections within the network were measured. The result showed that intervals up to a few hundred milliseconds could lead to distributed long-term modifications (both LTP and LTD) of polysynaptic pathways within the network [113]. While this study provides proof of principle for encoding long intervals through polysynaptic delays, does the natural neuronal network in the brain contain serially connected groups of synchronously firing neurons (synfire chain) [114] to encode long sequences and temporal intervals? The finding of sequential firing of neurons in the premotor area HVC of songbirds during each song syllable (which lasts >100 ms [115]) indicates the existence of synfire chains in the brain, the formation of which underlies song learning [116]. Spike sequences lasting for several seconds have also been observed in the mammalian neocortex [117].

In addition to STDP in polysynaptic networks, memories of time intervals on the order of seconds are also likely to involve other mechanisms. In a study in zebrafish larvae, Sumbre et al. [118] applied a unidirectional moving visual stimulus repeatedly at a fixed temporal interval (several seconds) and recorded the population activity of tectal neurons with Ca^{2+} imaging. They found that following cessation of the visual stimulation, sequential firing of the tectal neurons resembling that evoked by the moving stimulus reappeared at the same time interval as the conditioning stimuli, and this rhythmic reappearance lasted for up to 20 s. This post-conditioning spontaneous rhythmic firing of tectal neurons reflects short-term memory of a specific rhythm with time intervals of seconds. Interestingly, since the tectal activity did not persist continuously through the several seconds of interval, it is unlikely mediated by reverberation of activity in the polysynaptic network. The underlying circuit mechanism and the location of memory storage in this case remain unknown.

Remembering the association between events separated by intervals ranging from seconds to tens of seconds has been referred to as temporal associative learning, and has been found in recent studies to depend critically on entorhinal cortex (EC)–hippocampal circuits [119–121]. Pathway-selective inhibition and optogenetic activation of the EC–hippocampal circuit showed that layer II "island cells" and layer III neurons of the EC control temporal associative learning [119, 120]; in particular, persistent activity of medial EC layer III neurons may play a key role [121]. Recent studies have also shown the existence of hippocampal "time cells" that fire at particular moments in a temporally structured experience [122–124], suggesting a function that parallels that of place cells in spatial memory. The sequential activation of such time cells in the hippocampus may reflect temporally structured inputs from the cortex and other brain regions or alternatively firing chain generated within local hippocampal circuits by repetitive sequential activity-induced strengthening of synaptic connections.

Theories and models of sequence learning and interval timing have been proposed [125–127] but very few studies have directly addressed the circuit and synaptic mechanisms [128, 129]. Since the cellular and synaptic building blocks of memory are increasingly well characterized and the technology for manipulating specific cells is becoming available (see other contributions to this Forum), the storage mechanisms for sequence and interval information are now amenable to fruitful exploration. Understanding these mechanisms is a pre-requisite for further studies of circuit mechanisms underlying higher cognitive functions involving complex temporal information processing such as human language, the ultimate challenge to neuroscience.

Hippocampal sharp wave-ripple: a repetitive mechanism to support single trial learning

John Long and György Buzsáki

Memories are not "imprinted" immediately but evolve over time. Memory has many forms and supportive brain mechanisms. While learning complex skills and habits, such as walking elegantly in high-heel shoes or riding a unicycle, may require tens to thousands of repetitions, consciously remembering episodic information, such as recalling one's first date, often requires only a single trial. The brain can achieve such a feat by deploying a mechanism that repeats segments of the original episode subconsciously hundreds to thousands of times after the experience. Such repetitions occur during non-attentive, off-line states of brain operation in the form of hippocampal sharp wave-ripples (SPW-Rs). SPW-Rs operate as a time-compressing mechanism, which can

transfer information from the hippocampus to numerous regions of the neocortex. SPW-Rs also serve as a preconscious "mixer" of existing knowledge and recently acquired information. Thus, acquiring episodic memory is a two-stage process, initiated by a rapid encoding mechanism during attentive waking and followed by a protracted consolidation process during "off-line" states of the brain [130].

SPW-Rs are self-generated hippocampal patterns

The SPW-R complex is the most synchronous and phylogenetically preserved pattern in the mammalian brain, associated with enhanced transient excitability in the hippocampus and its partner structures. These super-synchronous bursts arise when the release of subcortical neuromodulators within the hippocampus is decreased—for example, during consummatory behaviors and slow wave sleep. A SPW-R is a superposition of two events, the sharp wave-related population burst (SPW), which emerges in the strongly recurrent system of the CA2/3 regions, and the fast ripple oscillation, which is dominant in the CA1 output circuit of the hippocampus. The synchronous discharge of CA3 pyramidal cells excites primarily the mid-apical dendrites of the CA1 region and the inward currents brought about by this transient depolarization process (40–150 ms) manifest extracellularly as the local field potential SPW. The response of target circuits to the strong SPW-related depolarization is a fast oscillatory balancing act between principal cells and perisomatic inhibitory interneurons, resulting in the LFP ripple and the phase-locked discharge of sequentially active neurons [130]. It is this waking history-dependent, sequential firing of a large fraction of hippocampal cells, and consequent recruitment of neocortical neurons, that makes SPW-Rs a candidate biomarker for memory consolidation.

Compressed replay of experience during SPW-Rs serves memory

The restructuring of hippocampal–cortical networks through synaptic plasticity is necessary for the formation of new episodic memories. Neurons participating in SPW-R events are organized to fire sequentially and the orderly structure of these events reflects a temporally compressed version of the sequential neuronal firing patterns observed in the waking animal [131, 132] (Fig. 5; "replay"). For example, the sequences of "place cells" [133] in a novel environment are formed from a combination of relatively fast-firing and slow-firing groups of pyramidal neurons. Conspicuously, the former neurons exhibit relatively unchanging temporal dynamics while the latter are highly plastic. The greater plasticity of slow-firing pyramidal neurons is evidenced by their greater gain in place specificity during maze exploration and their increased SPW-R-related recruitment during sleep following waking experience [134]. These sequences are not confined to the hippocampus–entorhinal system; the SPW-R output also brings about sequential activations in neocortical circuits, such as the prefrontal cortex [135], leading to systems-level consolidation of the memory trace. Even more direct evidence for the role of SPW-R sequences in memory consolidation is provided by close-loop truncation of SPW-Rs. Selective elimination of the hundreds of SPW-R-assisted replays of recently learned sequences during post-learning sleep prevented rats from becoming proficient at a spatial-reference memory task [136]. This supports the hypothesis that single-trial learning is made possible by repeated off-line SPW-R replays of the wake-experienced episode.

Memory and planning: retrospective and prospective roles for SPW- R

A particularly striking feature of SPW-Rs is that the neuronal sequences contained in them can propagate both forward and backward (Fig. 5) relative to waking experience. Before the beginning of a journey upon a familiar maze, the upcoming sequences are preplayed during SPW-Rs in a forward manner; that is, prior to a run, the CA1 pyramidal cells fire in a sequence consistent with the upcoming trajectory of the animal in the maze. At the end of the journey, again, the same neurons are reactivated during SPW-Rs but now in a reversed direction as the hippocampus recapitulates the landmarks passed by the animal, though in a time-compressed manner. These findings support the hypothesis that forward replay events play a role in "planning" upcoming trajectories [137]. This hypothesis is further supported by several recent experiments that demonstrate that the routes chosen by the animal in two-dimensional environments or between maze corridors can be predicted by the spike sequence content of SPW-Rs [138]. The subconscious route-priming role of SPW-Rs is also supported by other experiments, which show that truncating SPW-Rs prior to a choice results in a spatial working memory deficit [139]. Because the neurons participating in SPW-Rs sequences are drawn from a diverse pool of log-scale firing rate distributed neurons, with varying coding, biophysical, circuit, and plasticity properties, these events can transit a vast array of preexisting and new information to downstream cortical partners [134].

To date, memory mechanisms in experimental animals are typically studied in the framework of spatial navigation [133]. However, it is important to emphasize that mechanisms of memory and planning have evolved from mechanisms of navigation in the physical world. Therefore, neuronal algorithms underlying navigation in real and mental space, as well as place memory and episodic memory, are fundamentally the same [140].

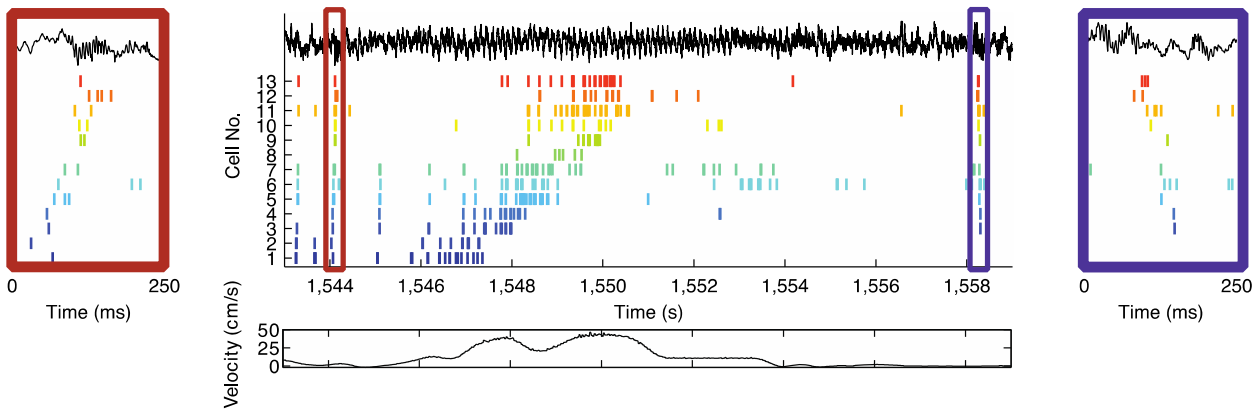


Fig. 5. Place cell sequences experienced during behavior (*middle panel*) are replayed in both forward (*left panel*) and reverse (*right panel*) direction during awake SPW-R. The rat is moving from left to right on a familiar track. Spike trains for place fields of 13 CA3 pyramidal cells (*color ticks*, spikes of individual neurons) on the track are shown before (forward replay; *left red box*), during (*middle*), and after (reverse replay; *right blue box*) a single traversal. The CA1 local field potential is shown on top (*black traces*) and the animal's velocity is shown below. Reproduced from [137]

Where is the study of Hebbian memory mechanisms going?

Charles F. Stevens

BMC Biology brought together some of the leaders in the learning and memory field to identify, from eight separate perspectives, some of the currently most important questions and approaches for understanding memory formation, consolidation, and retrieval. My job is to relate the current state of the field to how, in my opinion, it might evolve.

With rare, and highly valued, exceptions, a field develops organically from where it started and from the questions that grew out of each advance along the way. Although several of the contributions to this Forum relate key ideas from earlier eras (Cajal, Golgi, and Semon) to modern views, my starting point for the questions that drive learning and memory research will be Hebb's 1949 "firing - > wiring" proposal [1] for learning (cited in the introduction to this Forum and in four of the contributions).

Although Hebb's book was always well known to psychologists, it had essentially no impact on neurophysiology (as neuroscience was known in those days) for about a quarter of a century because the longest lasting form of synaptic plasticity known before 1973 (the date of Bliss and Lomo's LTP paper [5]) was post-tetanic potentiation (PTP) and this form of plasticity lasted, at most, only a few minutes. The only suggested brain mechanism for memory storage had been "reverberating circuits" and Von Neumann gave a compelling argument against the plausibility of this idea in his 1957 lectures at Yale on "The computer and the brain" [141]. Neurophysiologists were all very interested in the cellular basis of memory—everyone agreed it was a central problem—but no one had any idea how to study it.

As with any novel finding at odds with current knowledge and concepts, Bliss and Lomo's LTP paper [5] did not make an immediate splash. Indeed, during the next decade only a hand-full of researchers worked on LTP. But over the succeeding decades the study of LTP grew explosively to become a large and important sub-field of neuroscience, some would even say a preoccupation. As this Forum shows, we have learned an enormous amount about synaptic plasticity in the two-thirds of a century since Hebb's *The Organization of Behavior*. But where will we go now?

First, I need to confess that my record for predicting the future is abysmal: I have failed close to 100 % of the time. This time, however, I am trying a new strategy to see if I can move up from abysmal to, say, just terrible. This new approach is to see where the questions that drive our research have come from and then ask what we might have missed in asking these questions. Because of space constraints, I will focus on a single question: what might Hebbian learning be missing?

Although Behaviorism is pretty much dead [142], there is absolutely no question that reinforcement plays a key role in what behaviors an animal selects. And, as anyone who has experienced a traumatic event can tell you, the circumstances associated with the event are committed to an enduring memory (what were you doing when you heard about the 9/11/2001 terrorist attacks in the United States?). Of course, everyone is aware of this, yet the focus on Hebb's rule tends to neglect these other types of learning. Major exceptions to this are the climbing fibers in the cerebellum and the mossy fibers in the CA3 region of the hippocampus, where a single impulse or brief burst will fire the target neuron and is believed to potentiate the other synapses that are active at the same time. For both of these cases, Hebb's rule is sufficient to determine what needs to be learned.

So my specific prediction is that future scientists working on LTP/LTD (or STDP) will also include reward (positive or negative) related mechanisms in their investigations of synaptic plasticity at the molecular, cellular, circuit, and behavioral levels.

Where can we look to get guidance on wedding Hebb's rule to reinforcement learning? One of the simplest and best studied examples of alternative learning rules is the fruit fly olfactory system, where flies can learn to approach or avoid any odor if that odor is paired with reward (sugar water) or punishment (electric shock). A good example from Glenn Turner's laboratory [143] is an electrophysiological study showing how a specific dopamine neuron can decrease the synaptic strength of inputs into a specific mushroom body output neuron (MBON). The inputs into the MBON are synapses from 2000 mushroom body Kenyon cells and each odor activates a specific subpopulation of the Kenyon cells, about 100/2000. These 100 cells constitute a *tag* for that odor [144], a unique subpopulation of Kenyon cells whose firing stands in for the odor's combinatorial code generated by odorant receptor neurons in the fly's antennae. When the odor is paired with a shock, the specific dopamine neuron whose synapses overlap with the MBON's dendritic arbor and with the Kenyon cell's axonal arbor is activated. Firing of this dopamine neuron is what causes the Kenyon cell synapses to weaken and this change in synaptic strength is responsible for the avoidance of the odor associated with punishment.

Most scientists who study LTP/LTD/STDP use mammals as their experimental animals. It seems, however, that the dopamine system is, to a large degree, evolutionarily conserved even though the fly circuits are generally less complex than the corresponding ones in vertebrates. I am not aware of a review article that documents the insect/vertebrate parallels in the dopamine systems but the literature on the amygdala, a vertebrate center for valence learning, is vast [145] and can be compared to the large fly literature devoted to mushroom body learning and memory [146].

We can see over the next few years if my success for predicting future trends in science has improved.

Acknowledgements

Kelsey C. Martin

I thank S.L. Zipursky, S. Van Driesche, and J.T. Braslow for their comments, K. Otis for the photomicrograph in Fig. 2, and funding from R01NS045324.

Author details

[1]Institute of Neuroscience, Chinese Academy of Sciences, Shanghai, China. [2]RIKEN-MIT Center for Neural Circuit Genetics at the Picower Institute for Learning and Memory, Department of Biology and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]MPI of Neurobiology, Munich-Martinsried, Germany. [5]Department of Biological Chemistry and Department of Psychiatry and Biobehavioral Studies, David Geffen School of Medicine, BSRB 390B, 615 Charles E. Young Dr. South, University of California, Los Angeles, CA 90095, USA. [6]Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [7]The Neuroscience Institute, School of Medicine and Center for Neural Science, New York University, New York, NY 10016, USA. [8]Salk Institute for Biological Studies, Laboratory of Genetics, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA. [9]HHMI, Department of Molecular and Cell Biology, University of California, Berkeley, USA.

Published online: 19 May 2016

References

1. Hebb DO. The organization of behavior; a neuropsychological theory. New York: Wiley; 1949.
2. Katz LC, Shatz CJ. Synaptic activity and the construction of neural circuits. Science. 1996;274:1133–8.
3. Marr D. Simple memory: a theory for archicortex. Philos Trans R Soc Lond B Biol Sci. 1971;262(841):23–81.
4. Semon RW. The mneme. 1921.
5. Bliss TV, Lomo T. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. J Physiol. 1973;232(2):331–56.
6. Malinow R. Transmission between pairs of hippocampal slice neurons: quantal levels, oscillations, and LTP. Science. 1991;252(5006):722–4.
7. Markram H, Lübke J, Frotscher M, Sakmann B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science. 1997;275(5297):213–5.
8. Bi GQ, Poo MM. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci. 1998;18(24):10464–72.
9. Govindarajan A, Kelleher RJ, Tonegawa S. A clustered plasticity model of long-term memory engrams. Nat Rev Neurosci. 2006;7(7):575–83.
10. Mayford M, Siegelbaum SA, Kandel ER. Synapses and memory storage. Cold Spring Harb Perspect Biol. 2012;4(6):a005751.
11. Shi SH, Hayashi Y, Petralia RS, Zaman SH, Wenthold RJ, Svoboda K, et al. Rapid spine delivery and redistribution of AMPA receptors after synaptic NMDA receptor activation. Science. 1999;284(5421):1811–6.
12. Engert F, Bonhoeffer T. Dendritic spine changes associated with hippocampal long-term synaptic plasticity. Nature. 1999;399(6731):66–70.
13. Zhang F, Wang L-P, Boyden ES, Deisseroth K. Channelrhodopsin-2 and optical control of excitable cells. Nat Methods. 2006;3(10):785–92.
14. Liu X, Ramirez S, Pang PT, Puryear CB, Govindarajan A, Deisseroth K, et al. Optogenetic stimulation of a hippocampal engram activates fear memory recall. Nature. 2012;484(7394):381–5.
15. Ramirez S, Liu X, Lin P-A, Suh J, Pignatelli M, Redondo RL, et al. Creating a false memory in the hippocampus. Science. 2013;341(6144):387–91.
16. Ryan TJ, Roy DS, Pignatelli M, Arons A, Tonegawa S. Memory. Engram cells retain memory under retrograde amnesia. Science. 2015;348(6238):1007–13.
17. Tonegawa S, Pignatelli M, Roy DS, Ryan TJ. Memory engram storage and retrieval. Curr Opin Neurobiol. 2015;35:101–9.
18. Root CM, Denny CA, Hen R, Axel R. The participation of cortical amygdala in innate, odour-driven behaviour. Nature. 2014;515(7526):269–73.
19. Matsuzaki M, Honkura N, Ellis-Davies GCR, Kasai H. Structural basis of long-term potentiation in single dendritic spines. Nature. 2004;429(6993):761–6.
20. Le Bé J-V, Markram H. Spontaneous and evoked synaptic rewiring in the neonatal neocortex. Proc Natl Acad Sci U S A. 2006;103(35):13214–9.
21. Liao D, Hessler NA, Malinow R. Activation of postsynaptically silent synapses during pairing-induced LTP in CA1 region of hippocampal slice. Nature. 1995;375(6530):400–4.
22. Nabavi S, Fox R, Proulx CD, Lin JY, Tsien RY, Malinow R. Engineering a memory with LTD and LTP. Nature. 2014;511(7509):348–52.
23. Maletic-Savatic M, Malinow R, Svoboda K. Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. Science. 1999;283(5409):1923–7.
24. Crick F. Do dendritic spines twitch? Trends Neurosci. 1982;5:44–6.
25. Hofer SB, Mrsic-Flogel TD, Bonhoeffer T, Hubener M. Experience leaves a lasting structural trace in cortical circuits. Nature. 2009;457(7227):313–7.
26. Ebbinghaus H. Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie. Leipzig: Duncker & Humblot; 1885.
27. Hofer SB, Mrsic-Flogel TD, Bonhoeffer T, Hubener M. Prior experience enhances plasticity in adult visual cortex. Nat Neurosci. 2006;9(1):127–32.

28. Xu T, Yu X, Perlik AJ, Tobin WF, Zweig JA, Tennant K, et al. Rapid formation and selective stabilization of synapses for enduring motor memories. Nature. 2009;462(7275):915–9.
29. Yang G, Pan F, Gan WB. Stably maintained dendritic spines are associated with lifelong memories. Nature. 2009;462(7275):920–4.
30. Lai CS, Franke TF, Gan WB. Opposite effects of fear conditioning and extinction on dendritic spine remodelling. Nature. 2012;483(7387):87–91.
31. Moczulska KE, Tinter-Thiede J, Peter M, Ushakova L, Wernle T, Bathellier B, et al. Dynamics of dendritic spines in the mouse auditory cortex during memory formation and memory recall. Proc Natl Acad Sci U S A. 2013;110(45):18315–20.
32. Roberts TF, Tschida KA, Klein ME, Mooney R. Rapid spine stabilization and synaptic enhancement at the onset of behavioural learning. Nature. 2010;463(7283):948–52.
33. Hubener M, Bonhoeffer T. Searching for engrams. Neuron. 2010;67(3):363–71.
34. Hayashi-Takagi A, Yagishita S, Nakamura M, Shirai F, Wu YI, Loshbaugh AL, et al. Labelling and optical erasure of synaptic memory traces in the motor cortex. Nature. 2015;525(7569):333–8.
35. Davis HP, Squire LR. Protein synthesis and memory: a review. Psychol Bull. 1984;96(3):518–59.
36. Goelet P, Castellucci VF, Schacher S, Kandel ER. The long and the short of long-term memory–a molecular framework. Nature. 1986;322(6078):419–22.
37. Alberini CM. Transcription factors in long-term memory and synaptic plasticity. Physiol Rev. 2009;89(1):121–45.
38. Bliss TV, Collingridge GL. A synaptic model of memory: long-term potentiation in the hippocampus. Nature. 1993;361(6407):31–9.
39. Wang DO, Martin KC, Zukin RS. Spatially restricting gene expression by local translation at synapses. Trends Neurosci. 2010;33(4):173–82.
40. Cajigas IJ, Tushev G, Will TJ, tom Dieck S, Fuerst N, Schuman EM. The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. Neuron. 2012;74(3):453–66.
41. Miyashiro K, Dichter M, Eberwine J. On the nature and differential distribution of mRNAs in hippocampal neurites: implications for neuronal functioning. Proc Natl Acad Sci U S A. 1994;91(23):10800–4.
42. Poon MM, Choi SH, Jamieson CA, Geschwind DH, Martin KC. Identification of process-localized mRNAs from cultured rodent hippocampal neurons. J Neurosci. 2006;26(51):13390–9.
43. Aakalu G, Smith WB, Nguyen N, Jiang C, Schuman EM. Dynamic visualization of local protein synthesis in hippocampal neurons. Neuron. 2001;30(2):489–502.
44. Wang DO, Kim SM, Zhao Y, Hwang H, Miura SK, Sossin WS, et al. Synapse- and stimulus-specific local translation during long-term neuronal plasticity. Science. 2009;324(5934):1536–40.
45. Kang H, Schuman EM. A requirement for local protein synthesis in neurotrophin-induced hippocampal synaptic plasticity. Science. 1996;273(5280):1402–6.
46. Martin KC, Casadio A, Zhu H, Yaping E, Rose JC, Chen M, et al. Synapse-specific, long-term facilitation of aplysia sensory to motor synapses: a function for local protein synthesis in memory storage. Cell. 1997;91(7):927–38.
47. Kim S, Martin KC. Neuron-wide RNA transport combines with netrin-mediated local translation to spatially regulate the synaptic proteome. eLife. 2015;4.
48. Rajendran L, Bali J, Barr MM, Court FA, Kramer-Albers EM, Picou F, et al. Emerging roles of extracellular vesicles in the nervous system. J Neurosci. 2014;34(46):15482–9.
49. Zeisel A, Munoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, Jureus A, et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science. 2015;347(6226):1138–42.
50. Kasthuri N, Hayworth KJ, Berger DR, Schalek RL, Conchello JA, Knowles-Barley S, et al. Saturated reconstruction of a volume of neocortex. Cell. 2015;162(3):648–61.
51. Tsien RY. Very long-term memories may be stored in the pattern of holes in the perineuronal net. Proc Natl Acad Sci U S A. 2013;110(30):12456–61.
52. Kandel ER. The molecular biology of memory: cAMP, PKA, CRE, CREB-1, CREB-2, and CPEB. Mol Brain. 2012;5:14.
53. Kandel ER, Dudai Y, Mayford MR. The molecular and systems biology of memory. Cell. 2014;157(1):163–86.
54. Garner AR, Rowland DC, Hwang SY, Baumgaertel K, Roth BL, Kentros C, et al. Generation of a synthetic memory trace. Science. 2012;335(6075):1513–6.
55. Reijmers LG, Perkins BL, Matsuo N, Mayford M. Localization of a stable neural correlate of associative memory. Science. 2007;317(5842):1230–3.
56. Alarcon JM, Malleret G, Touzani K, Vronskaya S, Ishii S, Kandel ER, et al. Chromatin acetylation, memory, and LTP are impaired in CBP+/- mice: a model for the cognitive deficit in Rubinstein-Taybi syndrome and its amelioration. Neuron. 2004;42(6):947–59.
57. Korzus E, Rosenfeld MG, Mayford M. CBP histone acetyltransferase activity is a critical component of memory consolidation. Neuron. 2004;42(6):961–72.
58. Rubinstein JH, Taybi H. Broad thumbs and toes and facial abnormalities. a possible mental retardation syndrome. Am J Dis Child. 1963;105:588–608.
59. Fischer A, Sananbenesi F, Wang X, Dobbin M, Tsai LH. Recovery of learning and memory is associated with chromatin remodelling. Nature. 2007;447(7141):178–82.
60. Ricobaraza A, Cuadrado-Tejedor M, Marco S, Perez-Otano I, Garcia-Osta A. Phenylbutyrate rescues dendritic spine loss associated with memory deficits in a mouse model of Alzheimer disease. Hippocampus. 2012;22(5):1040–50.
61. Miller CA, Sweatt JD. Covalent modification of DNA regulates memory formation. Neuron. 2007;53(6):857–69.
62. Rudenko A, Dawlaty MM, Seo J, Cheng AW, Meng J, Le T, et al. Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. Neuron. 2013;79(6):1109–22.
63. Bredy TW, Wu H, Crego C, Zellhoefer J, Sun YE, Barad M. Histone modifications around individual BDNF gene promoters in prefrontal cortex are associated with extinction of conditioned fear. Learn Mem. 2007;14(4):268–76.
64. Haggarty SJ, Tsai LH. Probing the role of HDACs and mechanisms of chromatin-mediated neuroplasticity. Neurobiol Learn Mem. 2011;96(1):41–52.
65. Scandura JM, Roboz GJ, Moh M, Morawa E, Brenet F, Bose JR, et al. Phase 1 study of epigenetic priming with decitabine prior to standard induction chemotherapy for patients with AML. Blood. 2011;118(6):1472–80.
66. Graff J, Tsai LH. The potential of HDAC inhibitors as cognitive enhancers. Annu Rev Pharmacol Toxicol. 2013;53:311–30.
67. Chen S, Cai D, Pearce K, Sun PY, Roberts AC, Glanzman DL. Reinstatement of long-term memory following erasure of its behavioral and synaptic expression in Aplysia. eLife. 2014;3:e03896.
68. Frey U, Morris RG. Synaptic tagging and long-term potentiation. Nature. 1997;385(6616):533–6.
69. Madabhushi R, Gao F, Pfenning AR, Pan L, Yamakawa S, Seo J, et al. Activity-induced DNA breaks govern the expression of neuronal early-response genes. Cell. 2015;161(7):1592–605.
70. Tsai LH, Graff J. On the resilience of remote traumatic memories against exposure therapy-mediated attenuation. EMBO Rep. 2014;15(8):853–61.
71. Mullins C, Fishell G, Tsien RW. Unifying views of autism spectrum disorders: a consideration of autoregulatory feedback loops. Neuron. 2016. In press.
72. Bourne JN, Harris KM. Nanoscale analysis of structural synaptic plasticity. Curr Opin Neurobiol. 2012;22(3):372–82.
73. Yuste R, Bonhoeffer T. Morphological changes in dendritic spines associated with long-term synaptic plasticity. Annu Rev Neurosci. 2001;24:1071–89.
74. Bienenstock EL, Cooper LN, Munro PW. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. J Neurosci. 1982;2(1):32–48.
75. Deisseroth K, Bito H, Schulman H, Tsien RW. Synaptic plasticity: a molecular mechanism for metaplasticity. Curr Biol. 1995;5(12):1334–8.
76. Rabinowitch I, Segev I. Two opposing plasticity mechanisms pulling a single synapse. Trends Neurosci. 2008;31(8):377–83.
77. Shin SM, Zhang N, Hansen J, Gerges NZ, Pak DT, Sheng M, et al. GKAP orchestrates activity-dependent postsynaptic protein remodeling and homeostatic scaling. Nat Neurosci. 2012;15(12):1655–66.
78. Okuno H, Akashi K, Ishii Y, Yagishita-Kyo N, Suzuki K, Nonaka M, et al. Inverse synaptic tagging of inactive synapses via dynamic interaction of Arc/Arg3.1 with CaMKIIbeta. Cell. 2012;149(4):886–98.
79. Li L, Li Y, Tsien RW. Maintaining local receptor homeostasis along neuronal dendrites via endo-exocytic coupling. Poster abstract# 704.14/D18. San Diego: Society for Neuroscience.
80. Song S, Sjostrom PJ, Reigl M, Nelson S, Chklovskii DB. Highly nonrandom features of synaptic connectivity in local cortical circuits. PLoS Biol. 2005;3(3):e68.
81. Thiagarajan TC, Lindskog M, Malgaroli A, Tsien RW. LTP and adaptation to inactivity: overlapping mechanisms and implications for metaplasticity. Neuropharmacology. 2007;52(1):156–75.
82. Buzsaki G, Mizuseki K. The log-dynamic brain: how skewed distributions affect network operations. Nat Rev Neurosci. 2014;15(4):264–78.
83. Turrigiano GG, Nelson SB. Homeostatic plasticity in the developing nervous system. Nat Rev Neurosci. 2004;5(2):97–107.
84. Altman J, Das GD. Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. J Comp Neurol. 1965;124(3):319–35.
85. Eriksson PS, Perfilieva E, Bjork-Eriksson T, Alborn AM, Nordborg C, Peterson DA, et al. Neurogenesis in the adult human hippocampus. Nat Med. 1998;4(11):1313–7.
86. Spalding KL, Bergmann O, Alkass K, Bernard S, Salehpour M, Huttner HB, et al. Dynamics of hippocampal neurogenesis in adult humans. Cell. 2013;153(6):1219–27.

87. Clelland CD, Choi M, Romberg C, Clemenson Jr GD, Fragniere A, Tyers P, et al. A functional role for adult hippocampal neurogenesis in spatial pattern separation. Science. 2009;325(5937):210–3.
88. Sahay A, Scobie KN, Hill AS, O'Carroll CM, Kheirbek MA, Burghardt NS, et al. Increasing adult hippocampal neurogenesis is sufficient to improve pattern separation. Nature. 2011;472(7344):466–70.
89. Amaral DG, Scharfman HE, Lavenex P. The dentate gyrus: fundamental neuroanatomical organization (dentate gyrus for dummies). Prog Brain Res. 2007;163:3–22.
90. Chawla MK, Guzowski JF, Ramirez-Amaya V, Lipa P, Hoffman KL, Marriott LK, et al. Sparse, environmentally selective expression of Arc RNA in the upper blade of the rodent fascia dentata by brief spatial experience. Hippocampus. 2005;15(5):579–86.
91. Jung MW, McNaughton BL. Spatial selectivity of unit activity in the hippocampal granular layer. Hippocampus. 1993;3(2):165–82.
92. O'Reilly RC, McClelland JL. Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. Hippocampus. 1994;4(6):661–82.
93. Bakker A, Kirwan CB, Miller M, Stark CE. Pattern separation in the human hippocampal CA3 and dentate gyrus. Science. 2008;319(5870):1640–2.
94. Gilbert PE, Kesner RP, Lee I. Dissociating hippocampal subregions: double dissociation between dentate gyrus and CA1. Hippocampus. 2001;11(6):626–36.
95. Esposito MS, Piatti VC, Laplagne DA, Morgenstern NA, Ferrari CC, Pitossi FJ, et al. Neuronal differentiation in the adult hippocampus recapitulates embryonic development. J Neurosci. 2005;25(44):10074–86.
96. Schmidt-Hieber C, Jonas P, Bischofberger J. Enhanced synaptic plasticity in newly generated granule cells of the adult hippocampus. Nature. 2004;429(6988):184–7.
97. Li Y, Aimone JB, Xu X, Callaway EM, Gage FH. Development of GABAergic inputs controls the contribution of maturing neurons to the adult hippocampal network. Proc Natl Acad Sci U S A. 2012;109(11):4290–5.
98. Marin-Burgin A, Mongiat LA, Pardi MB, Schinder AF. Unique processing during a period of high excitation/inhibition balance in adult-born neurons. Science. 2012;335(6073):1238–42.
99. Ge S, Yang CH, Hsu KS, Ming GL, Song H. A critical period for enhanced synaptic plasticity in newly generated neurons of the adult brain. Neuron. 2007;54(4):559–66.
100. Nakashiba T, Cushman JD, Pelkey KA, Renaudineau S, Buhl DL, McHugh TJ, et al. Young dentate granule cells mediate pattern separation, whereas old granule cells facilitate pattern completion. Cell. 2012;149(1):188–201.
101. Ikrar T, Guo N, He K, Besnard A, Levinson S, Hill A, et al. Adult neurogenesis modifies excitability of the dentate gyrus. Front Neural Circuits. 2013;7:204.
102. Piatti VC, Ewell LA, Leutgeb JK. Neurogenesis in the dentate gyrus: carrying the message or dictating the tone. Front Neurosci. 2013;7:50.
103. Sahay A, Wilson DA, Hen R. Pattern separation: a common function for new neurons in hippocampus and olfactory bulb. Neuron. 2011;70(4):582–8.
104. Leutgeb JK, Leutgeb S, Moser MB, Moser EI. Pattern separation in the dentate gyrus and CA3 of the hippocampus. Science. 2007;315(5814):961–6.
105. Danielson NB, Kaifosh P, Zaremba JD, Lovett-Barron M, Tsai J, Denny CA, et al. Distinct contribution of adult-born hippocampal granule cells to context encoding. Neuron. 2016;90(1):101–12.
106. Ito M. Long-term depression. Annu Rev Neurosci. 1989;12:85–102.
107. Levy WB, Steward O. Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. Neuroscience. 1983;8(4):791–7.
108. Markram H, Lubke J, Frotscher M, Sakmann B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science. 1997;275(5297):213–5.
109. Bell CC, Han VZ, Sugawara Y, Grant K. Synaptic plasticity in a cerebellum-like structure depends on temporal order. Nature. 1997;387(6630):278–81.
110. Zhang LI, Tao HW, Holt CE, Harris WA, Poo M. A critical window for cooperation and competition among developing retinotectal synapses. Nature. 1998;395(6697):37–44.
111. Dan Y, Poo MM. Spike timing-dependent plasticity: from synapse to perception. Physiol Rev. 2006;86(3):1033–48.
112. Xu S, Jiang W, Poo MM, Dan Y. Activity recall in a visual cortical ensemble. Nat Neurosci. 2012;15(3):449–455;S441–2.
113. Bi G, Poo M. Distributed synaptic modification in neural networks induced by patterned stimulation. Nature. 1999;401(6755):792–6.
114. Abeles M. Local cortical circuits: an electrophysiological study. Berlin; New York: Springer-Verlag; 1982.
115. Long MA, Jin DZ, Fee MS. Support for a synaptic chain model of neuronal sequence generation. Nature. 2010;468(7322):394–9.
116. Okubo TS, Mackevicius EL, Payne HL, Lynch GF, Fee MS. Growth and splitting of neural sequences in songbird vocal development. Nature. 2015;528(7582):352–7.
117. Harvey CD, Coen P, Tank DW. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature. 2012;484(7392):62–8.
118. Sumbre G, Muto A, Baier H, Poo MM. Entrained rhythmic activities of neuronal ensembles as perceptual memory of time interval. Nature. 2008;456(7218):102–6.
119. Suh J, Rivest AJ, Nakashiba T, Tominaga T, Tonegawa S. Entorhinal cortex layer III input to the hippocampus is crucial for temporal association memory. Science. 2011;334:1415–20.
120. Kitamura T, Pignatelli M, Suh J, Kohara K, Yoshiki A, Abe K, et al. Island cells control temporal association memory. Science. 2014;343:896–901.
121. Kitamura T, Macdonald CJ, Tonegawa S. Entorhinal-hippocampal neuronal circuits bridge temporally discontiguous events. Learn Mem. 2015;22:438–43.
122. Eichenbaum H. Time cells in the hippocampus: a new dimension for mapping memories. Nat Rev Neurosci. 2014;15:732–44.
123. Manns JR, Howard MW, Eichenbaum H. Gradual changes in hippocampal activity support remembering the order of events. Neuron. 2007;56:530–40.
124. Pastalkova E, Itskov V, Amarasingham A, Buzsáki G. Internally generated cell assembly sequences in the rat hippocampus. Science. 2008;321:1322–7.
125. Buhusi CV, Meck WH. What makes us tick? Functional and neural mechanisms of interval timing. Nat Rev Neurosci. 2005;6(10):755–65.
126. Gallistel CR, Gibbon J. Time, rate, and conditioning. Psychol Rev. 2000;107(2):289–344.
127. Mauk MD, Buonomano DV. The neural basis of temporal processing. Annu Rev Neurosci. 2004;27:307–40.
128. Jazayeri M, Shadlen MN. A neural mechanism for sensing and reproducing a time interval. Curr Biol. 2015;25(20):2599–609.
129. Xu M, Zhang SY, Dan Y, Poo MM. Representation of interval timing by temporally scalable firing patterns in rat prefrontal cortex. Proc Natl Acad Sci U S A. 2014;111(1):480–5.
130. Buzsaki G. Hippocampal sharp wave-ripple: a cognitive biomarker for episodic memory and planning. Hippocampus. 2015;25(10):1073–188.
131. Dupret D, O'Neill J, Pleydell-Bouverie B, Csicsvari J. The reorganization and reactivation of hippocampal maps predict spatial memory performance. Nat Neurosci. 2010;13(8):995–1002.
132. Wilson MA, McNaughton BL. Reactivation of hippocampal ensemble memories during sleep. Science. 1994;265(5172):676–9.
133. O'Keefe J, Nadel L. The hippocampus as a cognitive map. Oxford: Oxford University Press; 1978.
134. Grosmark AD, Buzsaki G. Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. Science. 2016;351(6280):1440–3.
135. Peyrache A, Khamassi M, Benchenane K, Wiener SI, Battaglia FP. Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. Nat Neurosci. 2009;12(7):919–26.
136. Girardeau G, Benchenane K, Wiener SI, Buzsaki G, Zugaro MB. Selective suppression of hippocampal ripples impairs spatial memory. Nat Neurosci. 2009;12(10):1222–3.
137. Diba K, Buzsaki G. Forward and reverse hippocampal place-cell sequences during ripples. Nat Neurosci. 2007;10(10):1241–2.
138. Pfeiffer BE, Foster DJ. Hippocampal place-cell sequences depict future paths to remembered goals. Nature. 2013;497(7447):74–9.
139. Jadhav SP, Kemere C, German PW, Frank LM. Awake hippocampal sharp-wave ripples support spatial memory. Science. 2012;336(6087):1454–8.
140. Buzsaki G, Moser EI. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. Nat Neurosci. 2013;16(2):130–8.
141. Von Neumann J, Kurzweil R. The computer and the brain. Yale University Press; 2012.
142. Miller GA. The cognitive revolution: a historical perspective. Trends Cogn Sci. 2003;7(3):141–4.
143. Hige T, Aso Y, Modi MN, Rubin GM, Turner GC. Heterosynaptic plasticity underlies aversive olfactory learning in Drosophila. Neuron. 2015;88(5):985–98.
144. Stevens CF. What the fly's nose tells the fly's brain. Proc Natl Acad Sci U S A. 2015;112(30):9460–5.
145. Janak PH, Tye KM. From circuits to behaviour in the amygdala. Nature. 2015;517(7534):284–92.
146. Perisse E, Burke C, Huetteroth W, Waddell S. Shocking revelations and saccharin sweetness in the study of Drosophila olfactory memory. Curr Biol. 2013;23(17):R752–63.

Personality and Social Psychology Bulletin
http://psp.sagepub.com/

The Pull of the Past: When Do Habits Persist Despite Conflict With Motives?

David T. Neal, Wendy Wood, Mengju Wu and David Kurlander

Pers Soc Psychol Bull 2011 37: 1428 originally published online 22 August 2011
DOI: 10.1177/0146167211419863

The online version of this article can be found at:
http://psp.sagepub.com/content/37/11/1428

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
SPSP
Society for Personality and Social Psychology

Additional services and information for *Personality and Social Psychology Bulletin* can be found at:

Email Alerts: http://psp.sagepub.com/cgi/alerts

Subscriptions: http://psp.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://psp.sagepub.com/content/37/11/1428.refs.html

>> Version of Record - Sep 19, 2011

OnlineFirst Version of Record - Aug 22, 2011

What is This?

Downloaded from psp.sagepub.com at UNIV OF ILLINOIS URBANA on October 19, 2014

This version has been modified from the original version, which was included in the print issue of PSPB 37:11 (November 2011), to correct Figure 2.

In the print version, Figure 2 included an incorrect legend within the second graph. On p. 1434 of this corrected version, the incorrect legend has been removed from Figure 2.

Downloaded from psp.sagepub.com at UNIV OF ILLINOIS URBANA on October 19, 2014

Personality and Social Psychology Bulletin
37(11) 1428–1437
© 2011 by the Society for
Personality and Social Psychology, Inc
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146167211419863
http://pspb.sagepub.com

The Pull of the Past: When Do Habits Persist Despite Conflict With Motives?

David T. Neal[1], Wendy Wood[1], Mengju Wu[2], and David Kurlander[3]

Abstract

To identify the factors that disrupt and maintain habit performance, two field experiments tested the conditions under which people eat out of habit, leading them to resist motivational influences. Habitual popcorn eaters at a cinema were minimally influenced by their hunger or how much they liked the food, and they ate equal amounts of stale and fresh popcorn. Yet, mechanisms of automaticity influenced habit performance: Participants ate out of habit, regardless of freshness, only when currently in the context associated with past performance (i.e., a cinema; Study 1) and only when eating in a way that allowed them to automatically execute the response cued by that context (i.e., eating with their dominant hand; Study 2). Across all conditions, participants with weaker cinema-popcorn-eating habits ate because of motivations such as liking for the popcorn. The findings reveal how habits resist conflicting motives and provide insight into promising mechanisms of habit change.

Keywords

attitudes, automatic processes, health, self-regulation

Received August 27, 2010; revision accepted April 10, 2011

Habits develop when people give a response repeatedly in a particular context and thereby form associations in memory between the response and recurring context cues. Theories of automaticity illuminate the cognitive mechanisms behind habits. When people have a strong habit, perception of recurring context cues activates the response in memory and may additionally deactivate alternative responses (Danner, Aarts, & de Vries, 2008; Neal, Labrecque, Wood, & Lally, 2011). For example, when habitual runners were subliminally primed with words relating to the contexts in which they ran (e.g., *gym, forest*), the act of running became strongly accessible in thought (Neal et al., 2011). Suggesting that this context cuing process may be minimally influenced by motivations, subliminally priming participants' personal goals for running did not make running habits more accessible.

Habits initially are likely to form through goal pursuit—as people repeatedly perform actions that yield desired outcomes. However, once habits have developed, they are performed with only limited influence from supporting motivations (Triandis, 1977, 1979). Nonetheless, experimental tests of how these mechanisms play out with real-world behaviors have not yet been provided. The present research provides these tests by independently manipulating motivational and contextual factors across two field experiments with a socially significant behavior, habitual eating.

The basic question in our experiments is: What disrupts habit performance? The design pitted negative attitudes toward stale food against established habits to eat in a given context. If people continue to perform habits despite this substantial conflicting motive, then habits are resistant to changes in attitudes and goals. Although habits may not be easily influenced by motives, their dependence on context cues should make them vulnerable to disruptions in automaticity. That is, habits should not be *activated* automatically outside of their typical performance context and should not be *executed* automatically when responses are performed in novel ways (e.g., using the nondominant hand). By manipulating the factors that do and do not control habits, we can test whether strong habits become responsive to current motivations when automaticity is disrupted.

Evidence of the Automaticity Guiding Habitual Behavior

Social psychological experiments have largely focused on types of automaticity that flexibly accommodate to current

[1]University of Southern California, Los Angeles, CA, USA
[2]University of California, San Diego, CA, USA
[3]Duke University, Durham, NC, USA

Corresponding Authors:
David T. Neal or Wendy Wood, Department of Psychology, University of Southern California, Los Angeles, CA 90086
Email: david@empiricaresearch.com.au or wendy.wood@usc.edu

Downloaded from psp.sagepub.com at UNIV OF ILLINOIS URBANA on October 19, 2014

goals and attitudes (Kruglanski et al., 2002). Yet, research in neuroscience, animal learning, and correlational studies of everyday behavior has focused primarily on cued actions that are not influenced by current motives (reviewed in Wood & Neal, 2007, 2009).

In neuroscience research on the brain systems activated during behavioral tasks, performance of stimulus–response habits is localized in the basal ganglia, especially the putamen, whereas control of goal-directed actions is localized in other brain regions, often involving the prefrontal cortex (Balleine & O'Doherty, 2010; Yin & Knowlton, 2006). Thus, habit formation involves neural shifts from "largely evaluation-driven circuits to those engaged in performance" (Graybiel, 2008, p. 362). Furthermore, the different brain systems controlling habits and goal-directed behavior often compete in guiding action such that activity associated with goal-directed control may be actively suppressed during performance of habitual behavior (Balleine & O'Doherty, 2010; Poldrack et al., 2001). Animal learning research also differentiates habits from goal-directed actions, primarily in terms of behavioral reactions to rewards. Specifically, extended training at a task such as lever pressing for a food reward produces habitual, continued responding to the lever cue even after the reward is devalued (e.g., the animal is sated, the food has been associated with a toxin; Dickinson & Balleine, 1995).

The minimal influence of motivations on habit performance also is a central finding of correlational studies of everyday behavior, including *behavior prediction* (Triandis, 1977) and *habit discontinuity* (Verplanken, Walker, Davis, & Jurasek, 2008). Across a range of everyday behaviors—from bus travel, to fast food consumption, to exercise—motivational factors (e.g., attitudes, intentions, self-concept, attitude accessibility) predicted future performance for nonhabitual behaviors but had limited predictive power for strong habits (Ouellette & Wood, 1998). Furthermore, consistent with the idea that habits are cognitively represented as context–response associations, habit strength had these effects only when habits were assessed from frequent performance in stable contexts—the specific conditions that enable the formation of habit associations in memory (Aldrich, Montgomery, & Wood, in press; Danner et al., 2008). Discontinuity studies that assess naturally occurring context changes find that participants continue to perform habits with minimal influence from goals, but only so long as they continue to live in the same context (Verplanken et al., 2008; Wood, Tam, & Witt, 2005). When participants move house to a new location without cues to habit performance, they apparently are freed to act on their goals.

The findings from neuroscience, animal learning, and correlational research on everyday behavior suggest that, once developed, habits are performed with limited influence from motivations. It may be that strong habits are directly brought to mind by the context cues (e.g., settings, preceding actions) associated with prior performance. Then, through ideomotor processes, people may act on the accessible behavior in mind (Bargh, 1994). According to this view, for example, a person's habit to snack on crackers at night may come to be directly cued by sitting in front of the TV, and changes in motivational states of hunger and liking for food may have little influence on how much the person eats.

Attitudes and Goals Flexibly Guide Behavior

The insensitivity of habit performance to current motives outlined in the preceding research might seem surprising to social psychologists given the field's focus on the flexible operation of attitudes and goals. Although strong goals and attitudes are known to have enduring influence, they usually guide responding in a given context depending on the expected outcomes of an action and the value placed on those outcomes (Eagly & Chaiken, 1993). Thus, people act on strong food preferences in a given context depending on what food is available and how tasty it is. Even *automatic goal pursuit*—in which goals are activated and guide responding outside of awareness—yields "not a static behavioral response, but an automated *strategy* for dealing with the environment" (Bargh & Barndollar, 1996, p. 461, italics in original). Similarly, automatically activated attitudes often are malleable and context dependent (Dijksterhuis, Smith, van Baaren, & Wigboldus, 2005).

The flexible influence of implicit goals and attitudes is evident in research that opposes implicit against explicit motives. For example, Macrae and Johnston's (1998) participants responded to implicit priming of a helping goal by offering assistance to someone in need, except when the implicit helping goal was inconsistent with their explicit goal of being on time. In other research, participants carried out implementation intentions to study—a form of automatic goal pursuit in which people plan to perform a particular response upon encountering a particular cue—only when they held the explicit goal of studying (Sheeran, Webb, & Gollwitzer, 2005). In an experiment related to our present focus on eating habits, implicit goals relevant to thirst increased the amount that participants drank during a beverage taste test, but only when they were thirsty (Strahan, Spencer, & Zanna, 2002). Nonetheless, conflicts between explicit and implicit goals are not always resolved in favor of explicit ones (Shah & Kruglanski, 2002). Instead, these dispositions combine in dynamic ways to shape responding in goal pursuit.

Habits contrast with the flexible pattern of responding characteristic of motives. Unlike automatic goals and attitudes, performance of strong habits may not be dynamically responsive to current motives. Thus, a strong habit to eat a particular food in a given context may be insensitive to current hunger levels or the palatability of the food.

Downloaded from psp.sagepub.com at UNIV OF ILLINOIS URBANA on October 19, 2014

The Present Research

To test the factors that can alter habit performance, we conducted two field experiments varying the conditions under which people consume popcorn at a movie cinema. Some participants in our research had strong habits to eat popcorn at the cinema (i.e., a history of frequent popcorn consumption in that setting), whereas others had weaker habits. We chose to study eating behavior in part because it is a significant social problem and in part because eating provides a strong test of the sensitivity of habits to changes in motivations. People believe that their eating behavior is largely a motivated activity in response to the way food tastes (e.g., Vartanian, Herman, & Wansink, 2008). Indeed, 17 of 22 participants in a pilot study provided "tastes good" or a close synonym when asked why they ate popcorn at the cinema. Thus, in the present experiments, we manipulated attitudes toward the food through its palatability. For some participants, the popcorn was fresh, whereas for others it was 7 days old and stale (adapted from Wansink & Kim, 2005).

To test the effects of changes in motivations, we first assessed whether participants with strong and weak habits evaluated the popcorn in the same way. It is possible that strong-habit participants would report more favorable attitudes even toward stale popcorn. By assessing attitudes toward the specific popcorn served and not toward popcorn in general, we evaluated the immediate, proximal attitude relevant to participants' eating. We also assessed participants' reported hunger and, in Study 2, their perceptions of social norms. Our second test of the susceptibility of habits to shifts in motivations involved assessing how much popcorn participants ate. Specifically, we evaluated whether participants with strong habits ate more of the liked, fresh popcorn than the disliked, stale popcorn and ate more when they were hungry.

Even if habits are not responsive to changes in motives, they should be disrupted by challenges to habit cuing. In Study 1, participants were tested either in the context relevant to their habit (watching movies in a cinema) or in a novel context (watching music videos in a conference room). We anticipated that habits would be automatically brought to mind only in the cinema context. In Study 2, all participants ate in the habit-related context of a cinema, and we manipulated whether they ate in their typical way (with their dominant hand) or in a novel way (using their nondominant hand). Using one's nondominant hand has been shown to impede the automatic, smooth performance of habitual responses (Sakai, Kitaguchi, & Hikosaka, 2003), presumably because it brings the action under intentional control. Thus, eating habits should be automatically executed only when eating with the dominant hand.

If habits resist changes in motives but can be disrupted by changes in automatic cuing, then we anticipate the following three-way interaction: Participants at the cinema with strong popcorn habits should be guided by the action in mind and eat similar amounts of fresh and stale popcorn. In contrast, participants with weak or moderate habits should eat more of the fresh popcorn that they like. Alternatively, when automatic cuing is disrupted by eating in a novel context or eating in a cinema but in a novel manner, then even strong-habit participants are likely to fall back on their attitudes and consume more fresh than stale popcorn.

Study 1

Method

Design and participants. Study 1 used a Context (cinema vs. meeting room) × Food Attitudes (fresh vs. stale) between-participants design. Habit strength for eating popcorn at the movies was an additional, continuous predictor. For the cinema context, 98 participants (57 males) were recruited immediately before the showing of a regularly scheduled feature film at a campus cinema. For the meeting room context, 60 participants (35 males) were recruited through flyers and participated in a campus meeting room near the cinema. Participants were paid $6 for completing the study.[1]

Procedure

Participants in the cinema setting were told that the study examined personality differences in movie interests and that they would rate a series of movie trailers. Sessions were conducted in groups of around 15. Before entering the theater, participants rated their current feelings (hunger, time since last meal, several filler items) and were given a 591-ml bottle of water and a box of popcorn. Participants were randomly assigned to receive popcorn that was either fresh (popped 1 hr before the session) or stale (popped 7 days before the session). Each box was discretely numbered so that it could be matched to participants' survey responses. Boxes were weighed before and after distribution on a digital scale, with initial amount averaging 61.73 g (SD = 6.74).

Participants entered the theater with their popcorn and water and, to reduce potential social influence, sat as far as possible from other participants. The lights immediately dimmed, and a series of six movie trailers for unreleased films was shown, totaling 15 min of viewing time. All popcorn boxes and water containers were collected immediately after the final trailer. Participants then moved to the cinema foyer and completed the survey, which, to maintain the cover story, first assessed their interest in seeing each film and then included a personality inventory and the liking and habit strength measures (see below).

Participants in the meeting room condition followed a similar procedure with one critical modification. The study was conducted in a campus meeting room and involved watching and rating music videos, a novel setting not associated

Downloaded from psp.sagepub.com at UNIV OF ILLINOIS URBANA on October 19, 2014

Table 1. Mean Liking of Popcorn Received in Studies 1 and 2 as a Function of Experimental Condition

	Fresh		Stale	
Habit cuing manipulation	*M*	*SD*	*M*	*SD*
Study 1				
Cinema	3.51	(0.93)	2.60	(1.16)
Meeting room	3.60	(0.96)	3.03	(0.89)
Study 2				
Eating with dominant hand	4.18	(1.42)	3.10	(1.42)
Eating with nondominant hand	4.50	(1.31)	2.72	(1.36)

Higher numbers reflect more positive ratings of the popcorn on a 7-point scale ranging from *very bad* (1) to *very good* (7).

with prior popcorn eating. A pilot test ($N = 22$) confirmed that the music videos were equivalent to the movie trailers in terms of their interest value and attentional involvement ($ts < 1$). Instead of rating their interest in watching each movie, participants rated their interest in owning each music video. All other aspects of the design were identical to the cinema setting (e.g., lighting levels, popcorn boxes, time since last major meal, number and visibility of other participants).

Hunger ratings. Before watching the videos, participants rated their current hunger on a 5-point scale (anchors *not at all* to *extremely*).

Ratings of popcorn attitude. On a 7-point scale, participants rated their liking for the popcorn they received (anchors *very bad* to *very good*).

Habit strength. On a 7-point scale, participants indicated how frequently in the past they ate popcorn in movie theaters (anchors *always* to *never*, reverse scored).

Results and Discussion

Our primary hypotheses were tested via regression models with the following predictors: context (cinema vs. meeting room), popcorn freshness (stale vs. fresh), habit strength to eat popcorn in theaters, and all two- and three-way interactions. Gender was a control variable in all analyses.

Reported attitudes and hunger. To establish that the experiment was constructed appropriately to test our hypotheses, we first used the regression model to predict participants' liking of the popcorn they received. As expected, the model yielded only a main effect of the freshness manipulation, $B = 0.54$, $t(141) = 1.98$, $SE = 0.27$, $p = .05$, establishing that fresh popcorn was liked significantly more than stale. Importantly, this effect was not qualified by any interactions with context or habit strength.[2] Mean liking of fresh popcorn was close to the scale midpoint ($M = 3.56$, $SD = 1.23$), and mean liking of stale popcorn was below the midpoint ($M = 2.72$, $SD = 0.91$), demonstrating that participants decidedly disliked the stale popcorn (see Table 1). No other differences in liking or in self-reported hunger emerged across conditions. Thus, participants in the meeting room had similar levels of hunger to those in the cinema, $t(144) = 1.57$, $SE = 0.16$, $p = .117$. Furthermore, the comparable ratings of popcorn in both contexts suggests that participants in the meeting room did not pay any closer attention to the popcorn than those at the cinema, $t(144) = 1.30$, $SE = 0.18$, $p = .194$.

Factors influencing performance of eating habits. The percentage of popcorn participants consumed was then analyzed with the regression model. Significant main effects emerged for gender (men ate more), $B = 14.37$, $t(142) = 3.32$, $SE = 4.32$, $p = .001$; context, $B = 27.20$, $t(142) = 4.53$, $SE = 6.00$, $p = .001$; and freshness, $B = 15.96$, $t(142) = 2.27$, $SE = 7.05$, $p = .025$. The latter two effects were qualified by the predicted three-way interaction between habit strength, context, and freshness, $B = 11.36$, $t(142) = 2.04$, $SE = 5.56$, $p = .043$, and no other effects were significant. To test whether strong-habit participants' eating was activated automatically in the cinema context with little regard for freshness, we calculated simple regression slopes between percent of popcorn consumed and freshness of the popcorn separately for participants with weak, moderate, and strong consumption habits in the cinema (Cohen, Cohen, West, & Aiken, 2003; see Figure 1).

In the cinema context, participants with weak habits—who infrequently ate popcorn at cinemas—ate significantly less popcorn in the stale than in the fresh conditions, $B = –30.03$, $t(90) = 4.28$, $SE = 7.01$, $p = .001$. Those with moderate popcorn eating habits also ate less stale than fresh popcorn, $B = –19.87$, $t(90) = 4.94$, $SE = 4.02$, $p = .001$. Importantly, as predicted, those with strong popcorn eating habits ate the same percentage of stale as fresh popcorn when located in the cinema environment, $B = –9.72$, $t(90) = 1.41$, $SE = 6.89$, $p = .162$.

In the meeting room context, an effect of freshness was evident at every level of habit strength. Specifically, people ate (at least marginally) less stale than fresh popcorn, whether they had weak habits, $B = –11.24$, $t(90) = 1.86$, $SE = 6.04$, $p = .074$; moderate habits, $B = –17.21$, $t(90) = 4.09$, $SE = 4.21$, $p = 0.001$; or strong habits, $B = –23.17$, $t(90) = 3.03$, $SE = 7.65$, $p = .001$. Critically, this pattern demonstrates that when tested in a novel environment, participants with strong popcorn consumption habits ate less stale than fresh popcorn.

Habit performance did not depend on motives. We conducted separate analyses to evaluate directly whether habit performance depended on participants' hunger and liking for the popcorn. To evaluate hunger as a moderator, we included this factor as a predictor in the model. Although hungrier participants ate marginally more in general, $B = 4.01$, $t(142) = 1.7$, $SE = 2.36$, $p = .092$, the critical interaction between hunger and habit strength was not significant ($t < 1$), suggesting that habit effects did not depend on this motive. Even more importantly, after including hunger and the Hunger × Habit Strength interaction, the predicted three-way interaction maintained between habit strength, context, and freshness, $B = 12.27$, $t(142) = 2.17$, $SE = 5.66$, $p = .032$.

Downloaded from psp.sagepub.com at UNIV OF ILLINOIS URBANA on October 19, 2014

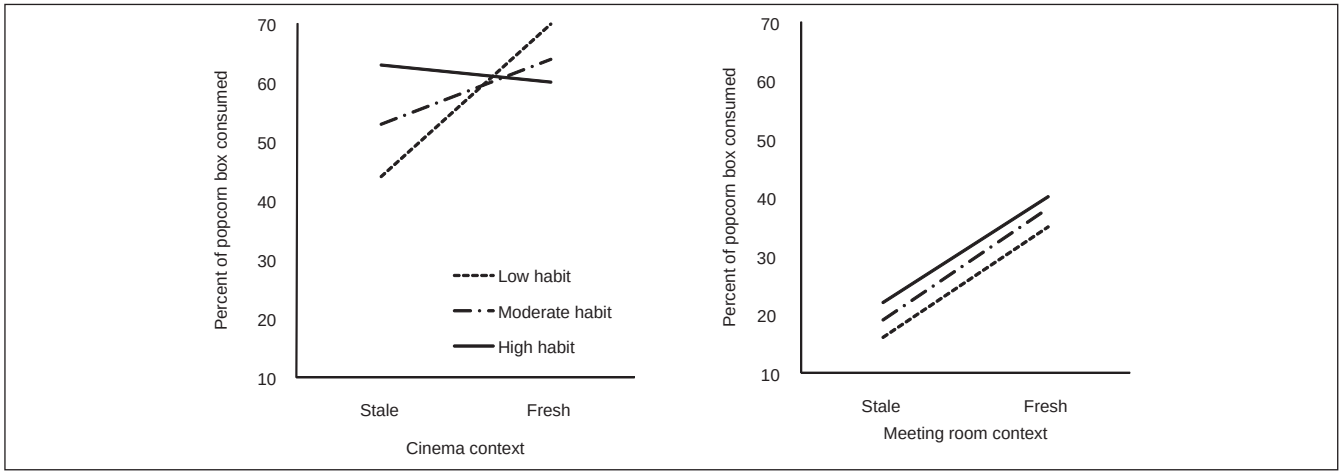


Figure 1. Percentage of popcorn eaten during 15 min of movie trailers in the cinema and 15 min of music videos in the meeting room (Study 1)

Simple slopes depict percent consumed of stale and fresh popcorn for those with strong (mean + 1 *SD*), moderate (mean), and weak (mean – 1 *SD*) habits for eating popcorn at cinemas.

We then conducted analyses evaluating liking as a moderator. Again, this motivation had a main effect, with participants eating more popcorn when they liked it more, $B = 8.32$, $t(142) = 3.95$, $SE = 2.11$, $p < .001$, but the interaction between liking and habit strength was not significant ($t < 1$), suggesting that habit effects did not depend on this attitude. Furthermore, the three-way interaction between habit strength, context, and freshness remained close to significant after including these effects in the model, $B = 9.69$, $t(142) = 1.80$, $SE = 5.35$, $p = .07$.

In general, the overall pattern suggests that habit performance resists shifts in attitudes but can be disrupted by shifts in habit cuing. It is important to note that strong-habit participants experienced the shift in motivation as a function of the food's freshness. They rated the fresh popcorn as more likable than the stale, and thus they did not value the popcorn more than those with weak habits. Because we assessed liking for the specific popcorn in the experiment and not popcorn in general, we evaluated the attitude toward the object most closely linked to consumption of this particular food. Although strong-habit participants experienced this motivational shift, their eating was not guided by it. That is, when in the cinema so that eating was activated automatically by the context, these participants consumed about 63% of the popcorn regardless of its palatability. Furthermore, tests of whether the habit effects depended on liking for the popcorn or current hunger revealed that these factors did not moderate the effects of habit strength on eating.

The eating behavior of strong-habit participants was, however, disrupted by features associated with automatic cuing. When in the meeting room watching movie videos, cinema-popcorn habits were not automatically activated, and strong-habit participants' behavior was brought under intentional control. Then, like the weak-habit participants in all conditions, they ate about half as much stale popcorn compared with fresh popcorn. Thus, participants' habitual eating was resistant to changes in motivational states but influenced by variations in habit automaticity.

Study 2

Study 2 replicated the test of insensitivity of habit performance to motivational states and tested an additional disruption of habit automaticity—whether the response could be repeated as in the past or had to be performed in a novel way. We anticipated that participants performing the response in a novel way would be unable to act automatically on any habitual response triggered by the cinema context. To test this idea, half of the participants were instructed to use their nondominant hand to eat. This disruption in the fluid execution of habitual eaters' consumption should not affect their judgments of the popcorn. Instead, the manipulation should hinder the automatic execution of the habitual response and bring behavior under intentional control—so that participants can be guided by their motivations and eat more fresh than stale popcorn. Thus, we tested whether participants with strong cinema-popcorn habits, when eating with their nondominant hand at the cinema, would be hindered in automatically repeating their past behavior and instead would become responsive to food freshness, consuming only popcorn they liked.

The study also assessed several potential alternative explanations for the predicted effects. To see whether the hand manipulation affected consumption by prompting people to attend more to the popcorn's freshness (and perhaps especially so for those with strong habits), we tested people's recall for specific details of the movie trailers. We reasoned that participants attending closely to the popcorn would, by definition, be attending less to the trailers and thus would recall fewer specific details. Also, based on prior work in

Downloaded from psp.sagepub.com at UNIV OF ILLINOIS URBANA on October 19, 2014

action identification theory (Vallacher & Wegner, 1989) and construal level theory (Trope & Liberman, 2010), we tested whether the hand manipulation encouraged people to construe popcorn consumption at a lower, more concrete level, thereby highlighting lower level, immediate outcomes such as taste. Finally, we assessed participants' perceived norms to eat popcorn and tested whether variation in normative beliefs could explain the effects of habit. Thus, Study 2 tested for changes in attention, construal level, and perceived norms as possible explanations for the predicted effects.

Method

Design and participants. The design of Study 2 was identical to Study 1 except that the context manipulation was replaced by a manipulation to eat with a particular hand. Thus, the design was a Hand Used to Eat (dominant vs. nondominant) × Food Freshness (fresh vs. stale) between-participants design. Eighty-nine participants (47 males) were paid $6 for completing the study.[3]

Procedure

Participants were recruited and participated just before the showing of a feature film at a campus cinema. Procedural details mirrored the cinema condition of Study 1, with the exception of the manipulation of the hand participants used to eat. For this manipulation, we designed a modified popcorn box with a vertically aligned handle on one side. Participants were instructed to slide one hand in-between the handle and the box and hold the box in that manner throughout the session. This instruction was embedded in several other methodological details and framed as necessary for experimental control. Testing sessions were randomly assigned such that, in half, participants were instructed to hold the box in their left (vs. right) hand.

In the follow-up survey, participants completed the Edinburgh Handedness Inventory (Oldfield, 1971). This enabled us to code participants according to whether they had eaten with their typical, dominant hand or their atypical, nondominant hand. Ambidextrous participants ($N = 4$) were coded as dominant hand eaters, but excluding this group did not alter the results reported below. At the study conclusion, participants were asked for an anonymous, honest estimate regarding the percentage of eating time in which they had actually used the instructed hand to eat. Estimates were made on an 11-point scale (*0% of the time* to *100% of the time*). Compliance with the hand manipulation was good. Ninety percent of the sample reported complying 100% of the time. Those who deviated did so an average of only 10% of the time. Excluding these participants or including compliance as a control variable did not alter the reported results.

Normative beliefs. To evaluate the extent to which habit performance depends on social norms, in the final survey participants rated on a 7-point scale the extent to which "other people important to them think that it's normal to eat popcorn at the cinema" (anchors *strongly disagree* to *strongly agree*). Because this item was not a close-to-significant predictor of consumption, it is not discussed further.

Attention to the trailers. To assess how much attention participants paid to the movie trailers, they answered nine difficult questions about the trailers (e.g., "What color did the main character in 'Whip It' dye her hair?"). The only close-to-significant effect in the analyses on factual recall was a trend for slightly worse recall among strong-habit participants, $B = –0.41$, $t(80) = 1.78$, $SE = 0.23$, $p = .07$. This pattern counters the possibility that strong-habit participants were not attending to the popcorn and were able to attend more closely to the movie trailers.

Level of construal of popcorn eating. Three questions assessed whether people construed the act of eating popcorn in the study at a higher, more abstract level versus a lower, more concrete level. Modeled after Vallacher and Wegner's (1989) research, each question involved choosing between a more concrete description of eating popcorn (i.e., chewing, holding a box, placing food in your mouth) and a more abstract one (i.e., part of the movie experience, sharing something with other people, a healthful snack). The sum of abstract choices (range = 0-3) served as the measure of construal level. Consistent with action identification theory and construal level theory, more habitual eaters preferred more abstract descriptions of popcorn consumption, $B = 0.14$, $t(81) = 2.51$, $SE = 0.07$, $p = .036$. However, the interaction between habit strength and hand used to eat did not significantly predict construal ($t > 1.2$). Thus, the predicted effects of the manipulation cannot be explained by changes in construal level.

Results and Discussion

The hypotheses were tested using a regression model with the following predictors: hand used to eat, popcorn freshness, habit strength to eat popcorn in theaters, and all two- and three-way interactions. Gender was a control variable in all analyses.

Reported attitudes and hunger. The model predicting participants' liking of the popcorn they received yielded only a main effect, reflecting that participants disliked the stale popcorn ($M = 3.43$, $SD = 1.59$) more than the fresh ($M = 4.35$, $SD = 1.59$), $B = 1.02$, $t(81) = 3.44$, $SE = 0.30$, $p = .001$. Importantly, this effect was not qualified by any interactions with context or habit strength.[4] As can be seen from the liking means in Table 1, no other differences emerged across conditions. Thus, mean liking did not vary with the hand used to eat, $t(85) = 0.78$, $SE = 0.33$, $p = .44$. Prestudy hunger levels also were identical among those eating with their dominant versus nondominant hand, $t(85) = 000$, $SE = 0.26$, $p = 1.00$.

Factors influencing performance of eating habits. The model predicting percentage of popcorn consumed yielded a main

Downloaded from psp.sagepub.com at UNIV OF ILLINOIS URBANA on October 19, 2014

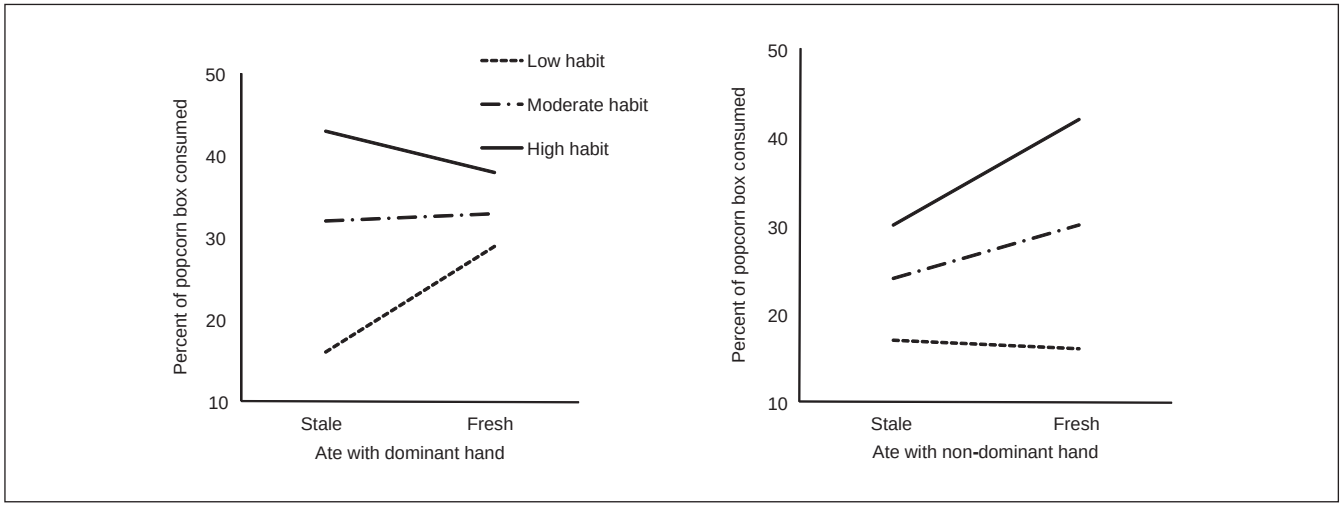


Figure 2. Percentage of popcorn eaten during 15 min of movie trailers in the cinema when participants were eating with their typical, dominant hand or atypical, nondominant hand (Study 2)
Simple slopes depict percent consumed of stale and fresh popcorn for those with strong (mean + 1 *SD*), moderate (mean), and weak (mean – 1 *SD*) habits for eating popcorn at cinemas.

effect of habit strength, $B = 6.74$, $t(81) = 2.55$, $SE = 2.64$, $p = .013$, and a marginal effect of hand used to eat, $B = 8.27$, $t(81) = 1.67$, $SE = 4.95$, $p = .094$. These main effects were qualified by the predicted three-way interaction between habit strength, hand used to eat, and freshness, $B = 8.52$, $t(81) = 2.22$, $SE = 3.84$, $p = .029$. To test whether strong habits were executed automatically— regardless of freshness— only when participants used their dominant hand, we calculated simple regression slopes between percent of popcorn consumed and freshness of the popcorn separately for participants with weak, moderate, and strong consumption habits as a function of the hand they used to eat (Cohen et al., 2003).

As predicted (see Figure 2), using the dominant versus nondominant hand to eat popcorn apparently enabled or disrupted habit cuing. That is, habitual eaters using their dominant hand automatically executed the response activated in the cinema context. The same amount of fresh and stale popcorn was eaten by participants with moderate habits, $B = –1.41$, $t(81) = 0.46$, $SE = 3.07$, $p = .647$, and strong habits, $B = 8.81$, $t(81) = 1.55$, $SE = 5.68$, $p = .13$. Only participants with weak habits ate significantly less stale than fresh popcorn when eating with the dominant hand, $B = –11.64$, $t(81) = 1.99$, $SE = 5.84$, $p = .053$.

Participants eating with their nondominant hand, in contrast, did not respond habitually to the cinema context and were influenced by popcorn freshness. Specifically, those with moderate habits ate significantly less stale than fresh popcorn, $B = –6.44$, $t(81) = 2.04$, $SE = 3.16$, $p = .048$, and those with strong habits ate marginally less, $B = –12.89$, $t(81) = 1.91$, $SE = 6.75$, $p = .063$. Unexpectedly, those with weak habits did not eat less stale than fresh popcorn when eating with their nondominant hand, $B = 0.02$, $t(81) = 0.01$, $SE = 2.00$, $p = .998$. However, inspection of the consumption rates for this group suggests a floor effect, such that eating with the nondominant hand was at such a low rate among these individuals that the stale popcorn could not induce further declines.

Habit performance did not depend on motives. As in Study1, we tested directly whether habit performance depended on participants' hunger and liking for popcorn. Although hungrier participants ate marginally more in general, $B = 2.72$, $t(79) = 1.76$, $SE = 1.55$, $p = .08$, the critical interaction between hunger and habit strength was not significant ($t < 1$), suggesting that habit effects did not depend on this motive. Even more importantly, after including hunger and the Hunger × Habit Strength interaction in the model, the predicted three-way interaction maintained between habit strength, context, and freshness, $B = 7.75$, $t(79) = 2.01$, $SE = 3.86$, $p = .048$.

In the analyses on liking as a moderator, this motivation yielded a main effect, with participants eating more popcorn if they liked it more, $B = 3.90$, $t(79) = 2.52$, $SE = 1.34$, $p = .014$. However, the interaction between liking and habit strength was not significant ($t < 1$), suggesting that the habit effects did not depend on this attitude. After adding these additional terms to the model, the critical three-way interaction between habit strength, context, and freshness dropped below significance, $B = 5.54$, $t(79) = 1.50$, $SE = 3.69$, $p = .13$. However, the low power to detect effects in this model may have contributed to this nonsignificant finding.

In general, just like Study 1, habitual cinema-popcorn eaters did not like the popcorn any more or less than nonhabitual eaters. Everyone disliked the stale popcorn and expressed neutral attitudes toward the fresh. It is also interesting that strong-habit participants expressed the same normative beliefs about eating popcorn in the cinema as weak-habit participants. Also, the slightly lower recall of the movie

Downloaded from psp.sagepub.com at UNIV OF ILLINOIS URBANA on October 19, 2014

trailers among strong-habit participants suggests that their attention was not unduly captured by these presentations, and thus they could attend to the quality of the popcorn as much as weak-habit participants. Finally, although strong-habit participants did favor more abstract descriptions of popcorn consumption, this tendency was not altered by the hand use manipulation, making changes in construal level an unlikely explanation for the results.

Despite the comparable subjective judgments of participants with strong and weak habits, striking differences emerged in their eating patterns. Strong-habit participants, when they could eat automatically with their dominant hand, repeated their past responses regardless of the palatability of the food. However, when strong-habit participants were forced to eat in an atypical way, their behavior was brought under intentional control, and—like the weak-habit participants in all conditions—they ate little of the stale popcorn.

General Discussion

Two field experiments tested the factors controlling a socially significant habit, eating behavior. In both experiments, participants' eating habits were resistant to changes in attitudes and goals. Strong habits persisted regardless of whether participants were hungry and whether the popcorn was fresh and palatable or stale and distasteful. Thus, when in the context associated with frequent past consumption, strong-habit participants rigidly carried out their past responses. Unlike the motivated types of automaticity studied most often in social psychology, habit responses were not sensitive to participants' current motivational states. Instead, strong-habit participants acted on habitual responses in memory to such an extent that they ate even food they disliked.

Habitual eating, although not influenced by motivations, was disrupted by factors that we anticipated would block the processes of habit automaticity. That is, participants in an environment not associated with past consumption (a meeting room) should not have the habitual response automatically brought to mind. Also, participants at the cinema eating in a novel way (with their nondominant hand) should not have been able to automatically carry out the response in mind. Our experiments thus directly manipulated the critical mechanisms that enable habit performance. When habit automaticity was apparently disrupted, participants' eating came under intentional control. That is, participants ate more when they were hungry and when they liked the popcorn—yielding greater consumption of fresh than stale popcorn. The findings thus echo prior research showing that people relying on their habits were less responsive to the outcomes of their behavior (Ji & Wood, 2007; Verplanken, Aarts, & van Knippenberg, 1997). By disrupting habitual eaters' automatic activation and execution of their past behavior, the present research reveals further that responsiveness to outcomes can be restored.

Importantly, the different eating patterns of strong- and weak-habit participants could not be explained by differences in the subjective evaluation of the popcorn. Habitual eaters recognized that the stale popcorn was not palatable, and they reported disliking it just as extremely as did nonhabitual eaters. It is striking, then, that these negative attitudes only curtailed habitual eaters' consumption when habit automaticity was disrupted. At the process level, the results do not appear to be due to changes in attention, social norms, or construal level across experimental conditions. Our check on recall of the details of the movie trailers in Study 2 revealed that the hand use manipulation did not change how much attention people paid to the popcorn, either as a main effect or through an interaction with habit strength. Also, all participants held similar norms to eat popcorn in the cinema. Finally, the hand use manipulation did not alter whether people thought about their popcorn consumption at a more abstract versus a more concrete level. Thus, it appears unlikely that the manipulation to disrupt habit automaticity worked through these alternative mechanisms.

The present study contributes to growing evidence that performance settings can serve as direct cues to habitual behavior. Habitual runners have been found to have strong cognitive associations between the location in which they typically trained and mental representations of running (Neal et al., 2011). Similarly, in behavior prediction studies, strong-habit participants tend to repeat responses independently of intentions when they are in familiar performance contexts (Ji & Wood, 2007). Based on this earlier research, we anticipated that the actual location of the theater might serve as a trigger to habit performance. Yet, the actual triggers might include a variety of aspects of the cinema setting and the evening show time. Because we had participants sit some distance from each other in the cinema, we doubt that movie companions were triggers for consumption in the present studies. Furthermore, because we used the same popcorn and the same bags in the cinema and meeting rooms, these features cannot have triggered habit performance. However, the exact nature of the triggering cues is a question for future research.

In general, our findings expand on work conducted by Wansink and colleagues in which the quantity of food people consume is influenced by simple manipulations of environmental cues such as plate and serving size (e.g., Sobal & Wansink, 2007; Wansink & Kim, 2005). Our results suggest that such environmental cues are likely to be most influential when people have developed habits to respond in particular ways to the cue. The current results also suggest a possible role for eating habits in Schachter's (1968) externality hypothesis, which stipulated that obese individuals are more driven by external cues and less by internal cues than are nonobese individuals (see also Wansink, Payne, & Chandon, 2007). Obesity has been linked to eating patterns such as nocturnal snacking that are repeated consistently at particular times of day (Berg et al., 2009). Thus, obese individuals may rely less than nonobese people on internal cues partly

Downloaded from psp.sagepub.com at UNIV OF ILLINOIS URBANA on October 19, 2014

because their eating behavior is more habitual, leading them to respond as cued by environments of past consumption.

Finally, our findings suggest ways to gain traction on the difficult problem of habit change (see Rothman, Sheeran, & Wood, 2009; Verplanken & Wood, 2006). Consistent with the idea that habits are a form of non-goal-dependent automaticity (see Bargh, 1994; Moors & De Houwer, 2006), habits may not be especially sensitive to alterations in motivational states but instead may best be disrupted by changes in triggering contexts. Habit change may thus require impeding habit activation or interrupting fluid habit execution. Although our findings suggest that both avenues are effective, it is not always possible for dieters to avoid or alter the environments in which they typically overeat (see Quinn, Pascoe, Wood, & Neal, 2010). More feasible, perhaps, is for dieters to actively disrupt the execution of the activated eating sequence by simple manipulations such as eating with the nondominant hand and, in so doing, bring their eating under their personal control.

Authors' Note

The authors are grateful to Meredith McAdams, Connie Wang, and Fade Eadeh for their help with data collection.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. In Study 1, 16 additional participants (7 in the cinema context, 9 in the meeting room context) were excluded from the analysis because they either declined the popcorn or did not eat any and thus were not exposed to the freshness manipulation.
2. To further test for possible effects of habit strength on liking, we created high- and low-habit groups via a median split of the habit variable. As with the continuous measure, strong-habit participants liked the fresh popcorn ($M = 3.66$, $SD = 0.77$) significantly more than the stale popcorn ($M = 2.70$, $SD = 1.11$), and weak-habit participants also liked the fresh popcorn ($M = 3.33$, $SD = 1.14$) more than the stale popcorn ($M = 2.83$, $SD = 1.05$). An ANOVA on this dichotomous measure of habit strength yielded only a main effect of freshness, $F(1, 146) = 18.58$, $p < .001$, and the Habit × Freshness interaction term was not significant ($F < 2$).
3. In Study 2, 15 additional participants (8 in the right-hand condition, 7 in the left-hand condition) were excluded because they either declined the popcorn box or did not eat any popcorn and thus were not exposed to the freshness manipulation.
4. As in Study 1, we used a median split of habit strength to explore potential differences in liking. Again, strong-habit participants liked the fresh popcorn ($M = 4.38$, $SD = 1.76$) significantly more than the stale popcorn ($M = 3.00$, $SD = 1.60$), and weak-habit participants also liked the fresh popcorn ($M = 4.25$, $SD = 1.12$) more than the stale popcorn ($M = 2.84$, $SD = 1.18$). An ANOVA on this dichotomous measure of habit strength yielded only a main effect of freshness, $F(1, 83) = 19.16$, $p < .001$, and the habit and Habit × Freshness interaction terms were not significant (Fs > 1).

References

Aldrich, J. H., Montgomery, J., & Wood, W. (in press). Turnout as a habit. *Political Behavior*.

Balleine B. W., & O'Doherty, J .P. (2010). Human and rodent homologies in motor control: Cortico-striatal determinants of goal-directed and habitual action. *Neuropsychopharmacology, 35*, 48-69. doi:10.1038/npp.2009.131

Bargh, J. A. (1994). The Four Horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In R. S. Wyer Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1-40). Hillsdale, NJ: Erlbaum.

Bargh, J. A., & Barndollar, K. (1996). Automaticity in action: The unconscious as repository of chronic goals and motives. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action* (pp. 457-471). New York, NY: Guilford.

Berg, C., Lappas, G., Wolk, A., Strandhagen, E., Tóren, K., Rosengren, A., . . . & Lissner, L. (2009). Eating patterns and portion size associated with obesity in a Swedish population. *Appetite, 52*, 21-26. doi:10.1016/j.appet.2008.07.008

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavior sciences.* Mahwah, NJ: Erlbaum.

Danner, U. N., Aarts, H., & de Vries, N. K. (2008). Habit vs. intention in the prediction of future behavior: The role of frequency, context stability and mental accessibility of past behavior. *British Journal of Social Psychology, 47*, 245-265. doi:10.1348/014466607X230876

Dickinson, A., & Balleine, B. (1995). Motivational control of instrumental action. *Current Directions in Psychological Science, 4*, 162-167. doi:10.1111/1467-8721.ep11512272

Dijksterhuis, A., Smith, P. K., van Baaren, R. B., & Wigboldus, D. H. J. (2005). The unconscious consumer: Effects of environment of consumer choice. *Journal of Consumer Psychology, 15*, 193-202. doi:10.1207/s15327663jcp1503_3

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt, Brace & Jovanovich.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience, 31*, 359-387. doi:10.1146/annurev.neuro.29.051605.112851

Ji, M., & Wood, W. (2007). Purchase and consumption habits: Not necessarily what you intend. *Journal of Consumer Psychology, 17*, 261-276. doi:10.1016/S1057-7408(07)70037-2

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Chun, W. Y., & Sleeth-Keppler, D. (2002). A theory of goal systems. *Advances in Experimental Social Psychology, 34*, 331-378.

Macrae, C. N., & Johnston, L. (1998). Help, I need somebody: Automatic action and inaction. *Social Cognition, 16*, 400-417.

Downloaded from psp.sagepub.com at UNIV OF ILLINOIS URBANA on October 19, 2014

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132,* 297-326. doi:10.1037/0033-2909.132.2.297

Neal, D. T., Labrecque, J., Wood, W., & Lally, P. (2011). *How do habits guide behavior? Perceived and actual triggers of habits in daily life.* Manuscript under review, University of Southern California, Los Angeles.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*, 97-113. doi:10.1016/0028-3932(71)90067-4

Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin, 124*, 54−74. doi: 10.1037/0033-2909.124.1.54

Poldrack, R. A., Clark, J., Pa're-Blagoev, J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414*, 546−550. doi:10.1038/35107080.

Quinn, J. M., Pascoe, A. M., Wood, W., & Neal, D. T. (2010). Can't control yourself? Monitor those bad habits. *Personality and Social Psychology Bulletin, 36*, 499-511. doi:10.1177/0146167209360665

Rothman, A. J., Sheeran, P., & Wood, W. (2009). Reflective and automatic processes in the initiation and maintenance of food choices. *Annals of Behavioral Medicine, 28*(Suppl.), 4-17. doi:10.1007/s12160-009-9118-3

Sakai, K. S., Kitaguchi, K., & Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. *Experimental Brain Research, 152,* 229-242. doi:10.1007/s00221-003-1548-8

Schachter, S. (1968). Obesity and eating. *Science, 161,* 751-756. doi:10.1126/science.161.3843.751

Shah, J. Y., & Kruglanski, A. W. (2002). Priming against your will: How goal pursuit is affected by accessible alternatives. *Journal of Experimental Social Psychology, 38,* 368-383. doi:10.1016/S0022-1031(02)00005-7

Sheeran, P., Webb, T. L., & Gollwitzer, P. M. (2005). The interplay between goal intentions and implementation intentions. *Personality and Social Psychology Bulletin, 31*, 87-98. doi:10.1177/0146167204271308

Sobal, J., & Wansink, B. (2007). Kitchenscapes, tablescapes, platescapes, and foodscapes: Influences of microscale built environments on food intake. *Environment and Behavior, 39*, 124-142. doi:10.1177/0013916506295574

Strahan, E., Spencer, S. J., & Zanna, M. P. (2002). Subliminal priming and persuasion: Striking while the iron is hot. *Journal of Experimental Social Psychology, 38*, 556-568. doi:10.1016/S0022-1031(02)00502-4

Triandis, H. C. (1977). *Interpersonal behavior*. Monterey, CA: Brooks/Cole.

Triandis, H. C. (1979). Values, attitudes, and interpersonal behavior. *Nebraska Symposium on Motivation, 27,* 195-259.

Trope. Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117,* 440-463. doi:10.1037/a0018963.

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology, 57,* 660-671.

Vartanian, L. R., Herman, C. P., & Wansink, B. (2008). Are we aware of the external factors that influence our food intake? *Health Psychology, 27,* 533-538. doi:10.1037/0278-6133.27.5.533

Verplanken, B., Aarts, H., & van Knippenberg, A. (1997). Habit, information acquisition, and the process of making travel mode choices. *European Journal of Social Psychology, 727,* 539-560. doi:10.1002/(SICI)1099-0992(199709/10)27:5<539::AID-EJSP831>3.0.CO;2-A

Verplanken, B., Walker, I., Davis, A., & Jurasek, M. (2008). Context change and travel mode choice: Combining the habit discontinuity and self-activation. *Journal of Environmental Psychology, 28,* 121-127. doi:10.1016/j.jenvp.2007.10.005

Verplanken, B., & Wood, W. (2006). Interventions to break and create consumer habits. *Journal of Public Policy and Marketing, 25,* 90-103. doi:10.1509/jppm.25.1.90

Wansink, B., & Kim, J. (2005). Bad popcorn in big buckets: Portion size can influence intake as much as taste. *Journal of Nutrition Education and Behavior, 37,* 242-245. doi: 10.1016/S1499-4046(06)60278-9

Wansink, B., Payne, C. R., & Chandon, P. (2007). Internal and external cues of meal cessation: The French paradox redux? *Obesity, 15,* 2920-2924. doi:10.1038/oby.2007.348

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review, 114,* 843-863. doi:10.1037/0033-295X.114.4.843

Wood, W., & Neal, D. T. (2009). The habitual consumer. *Journal of Consumer Psychology, 19*, 579-592. doi:10.1016/j.jcps.2009.08.003

Wood, W., Tam, L., & Witt, M. G. (2005). Changing circumstances, disrupting habits. *Journal of Personality and Social Psychology, 88*, 918-933. doi:10.1037/0022-3514.88.6.918

Yin, H., & Knowlton, B. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience, 7*, 464-476. doi:10.1038/nrn1919

Downloaded from psp.sagepub.com at UNIV OF ILLINOIS URBANA on October 19, 2014